PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : G06F 17/30 — A2

(11) International Publication Number: WO 00/57309

(43) International Publication Date: 28 September 2000 (28.09.00)

(21) International Application Number: PCT/US00/07474

(22) International Filing Date: 20 March 2000 (20.03.00)

(30) Priority Data:
09/272,814 — 19 March 1999 (19.03.99) — US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 09/272,814 (CON)
Filed on 19 March 1999 (19.03.99)

(71) Applicant *(for all designated States except US)*: STRUCTURAL BIOINFORMATICS, INC. [US/US]; 10929 Technology Place, San Diego, CA 92127 (US).

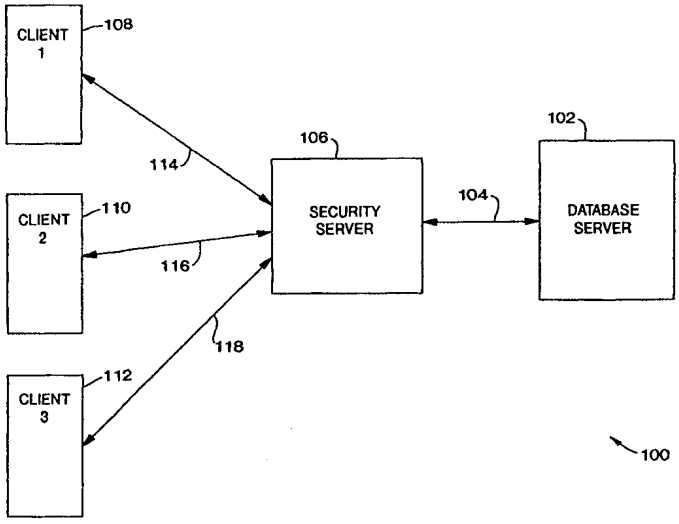
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: RAMNARAYAN, Kalyanaraman [IN/US]; 11674 Springside Road, San Diego, CA 92128 (US). VESSAL, Behnam [IR/US]; 657 Santa Camelia, Solana Beach, CA 92075 (US). KOTTALAM, Jeyapandian [IN/US]; 751 Avenida Amigo, San Marcos, CA 92069 (US). FISHER, Cindy, L. [US/US]; 3905 Carta De Plata, San Clemente, CA 92673 (US). MOEZZI, Saied [US/US]; 10420 Caminito Alvarez, San Diego, CA 92126 (US). PRABHAKARAN, Muthuchidambaram [US/US]; 13267 Treecrest Street, Poway, CA 92064 (US).

(74) Agents: SEIDMAN, Stephanie, L. et al.; Heller Ehrman White & McAuliffe LLP, 4250 Executive Square, Suite 700, La Jolla, CA 92037 (US).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DATABASE AND INTERFACE FOR 3–DIMENSIONAL MOLECULAR STRUCTURE VISUALIZATION AND ANALYSIS



(57) Abstract

A molecular structure database system collects multiple data files relating to the same molecule in the same subdirectory, and provides an interface to access all of the collected files from the same molecule using a graphical user interface (GUI) program. The collected files can comprise a variety of information and computer file formats, depending on the type of information to be conveyed to users of the database. A user communicates over a shared network with a secure file server that controls access to the collected files, and the interface to the collected files is provided by a GUI program or client applets. This provides a convenient means of searching molecular structure data for characteristics of interest. Data searching, file viewing, and investigation of multiple representations of molecular structures can be carried out from within a single viewing program.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

DATABASE AND INTERFACE FOR 3-DIMENSIONAL MOLECULAR STRUCTURE VISUALIZATION AND ANALYSIS

CROSS-REFERENCE TO RELATED APPLICATION

For purposes of International patent applications, this application claims benefit of priority from U.S. Non-provisional patent application Serial No. 09/272,814 entitled "Database and Interface for 3-Dimensional Molecular Structure Visualization and Analysis" filed March 19, 1999. For purposes of any U.S. application, priority is claimed from U.S. Non-provisional patent application Serial No. 09/272,814 entitled "Database and Interface for 3-Dimensional Molecular Structure Visualization and Analysis" filed March 19, 1999. Where permitted, this application incorporates by reference the referenced pending U.S. patent application in its entirety for all purposes.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates generally to database access and, more particularly, to interfaces that control access to collections of data.

2. Structure-Based Drug Design

Recent advances in molecular biology, such as the discovery and identification of large numbers of novel gene sequences encoded in the genomes of humans, other mammals and infectious disease agents, have contributed to the identification of a large number of target proteins and other biological macromolecules. Within the decade, up

to 150,000 fully sequenced human genes and an estimated 400,000 genes from other species will be available.

Based on information derived from these sequences, as well as from experimental methods, such as X-ray crystallography or NMR, or protein structure determination methodologies, such as homology modeling or *ab initio* structure prediction, three-dimensional (3-D) molecular structures of enzymes, ligands or target receptors, such as protein or other macromolecular receptors, are being determined at increasing rates. 3-D molecular structure is known to be related to biological function. By employing structure-based drug discovery methodologies, including structure-directed combinatorial or molecular diversity screening and computational screening using molecular similarity or computational docking algorithms, it is possible to derive knowledge of 3-D structure of biomolecules, which is useful, for example, in facilitating the identification of pharmacophores and the design of biologically-active compounds, such as small molecule agonists or antagonists.

As the number of available biomolecular structures increases, there is an increasing need for capabilities, such as databases, for organizing and providing access to the structures to make them available for use in structure-based discovery and design.

3. 3-Dimensional Molecular Structures

3-D molecular structures of proteins or other molecules are represented as sets of atomic coordinates that describe the spatial arrangement and intramolecular connectivity of the atoms in the molecule. For example, a standard format of representing macromolecular structural data is specified by the Protein Data Bank (PDB) format. The

PDB format organizes molecular structure data by representing atoms according to their atom types and atomic coordinates. In addition to atomic coordinate data, a PDB format data file can include information on structural attributes or reactivity of the molecule, such as active sites or secondary structure attributes. The molecular structure data in a PDB data file is provided as a simple alphanumeric representation of the atomic coordinates for a single protein or other macromolecule. Thus, it is stored as a text file using standard ASCII display codes.

The PDB is a depository of molecular structure data for biological macromolecules, that is, researchers are invited to add molecular structure data in the PDB format to the collection currently maintained at the Brookhaven National Laboratory (Bernstein et al., "The protein data bank: a computer-based archival file for macromolecule structures", J. Mol. Biol., 112, 535-542 (1977)). To facilitate maximum accessibility to the data contained in the PDB, the data files are made available over the Internet and can be retrieved and viewed by a user. Unfortunately, the PDB collection of data files is simply an archive comprising a series of molecular data files stored serially, as the files are deposited by researchers. That is, the PDB data files are flat files. Thus, there is no logical structure or interrelationship among the data files or between the records of different files. If a user is interested in viewing the atomic structure of a particular protein, for example, then the user must conduct a text-based search of the alphanumeric data in each file, one after the other, until the desired protein name or structure, or other associated information, is located. Text searching through the molecular data files at a depository such as the PDB might not effectively identify the

desired molecular structural information. Thus, it would be advantageous if molecular structure data files could be conformed to a logical database design, to permit more convenient and efficient searching for data records.

The coordinates of molecular structures can be read or downloaded into a molecular graphics program for visualization and manipulation and structural analysis. For example, a researcher can textually search the PDB data file to locate a protein of interest, and then can import the located file into a viewer program to obtain a visual representation of the protein. Several such graphics packages are known and are readily available commercially. Molecular graphics programs read in atomic coordinate data, such as is contained in a PDB-format file, and perform calculations to construct a three-dimensional representation of the molecule. Additionally, data beyond simple molecular structure, such as molecular shape analysis, energetic or strain analysis, active or reactive sites, variants or properties such as electrostatic potential or hydrophobicity, among others, may be associated with a given molecular structure. It would also be desirable to provide access to this information, as well. Currently, many of these data files are generated and must be viewed using different programs. To date, there is no software package available that integrates a molecular graphics program with a relational database to permit navigation among related molecular structures stored in a database and visualization analysis of molecular structures and their related properties within a single user interface. Thus, it would be further advantageous to integrate a molecular structure database with molecular visualization and analysis tools, as navigating among the

different databases and viewing programs can be time consuming and can be confusing, as each program may have a different scale or look and feel.

4. Proprietary Data Issues

Recently, techniques such as homology modeling and other structure prediction methodologies have been used to generate models of enzymes, ligands or target receptors, such as protein or other macromolecular receptors, structures whose structures have not yet been determined experimentally. The databases described herein can provide access to these structures to researchers and others, such as clinicians or educators. The databases can be stored on large networks, such as the Internet, which provide a convenient means of data dissemination, but are most well-known for providing unrestricted access to data uploaded to Internet servers.

In some cases, however, it might be desirable for the proprietor of a molecular structure database to carefully control access to their respective collections of molecular structure data files. For example, such data files may be obtained as a result of proprietary structure prediction algorithms or methods. In such cases, access to such information might be limited to only certain structures or families of structures.

Thus, it would be additionally advantageous to provide an integrated system to permit controlled access to a database of molecular structures and related properties within a single user interface.

SUMMARY OF THE INVENTION

Described herein is a database and interface for access to 3-D molecular structures and associated properties, which can be used to facilitate the design of potential new therapeutics. The interface also provides access to other structure-based drug discovery tools and to other databases, such as databases of chemical structures, including fine chemical or combinatorial libraries, for use in structure-focused high-throughput screening, as well as to a host of public domain databases and bioinformatics sites.

The invention provides a relational database that collects multiple data files relating to the same molecular structure in the same subdirectory, and provides an interface to access all of the collected files from the same structure using the same user interface program. The collected files can comprise a variety of information and computer file formats, depending on the type of information to be conveyed to users of the database. In accordance with the invention, a user communicates over a public shared network, such as the Internet, or over a controlled private network, such as an intranet, with a secure server that controls access to the collected files, and the interface to the collected files is provided by a graphical user interface. In this way, the invention provides a convenient means of searching molecular structure data for characteristics of interest. The invention also permits data searching, file viewing, and investigation of multiple representations of molecular structures from within a single viewing program.

In one aspect of the invention, the data files are made available over a wide area shared network, such as the Internet, and the graphical user interface (GUI) used for viewing the data files is a standard Internet web browser program, such as the web

browser products by Netscape Communications, Inc. and Microsoft Corporation. Such browser products readily import and provide views of files having a wide variety of formats that contain alphanumeric, video, and audio data. In another aspect of the invention, the GUI is provided with a platform independent programming environment using client applets, such as Enterprise Java Beans (EJB). A security server is preferably located between the user browser program at a network client machine controls access to the database, which is housed at a file server connected to the security server. Before a user gains access to the database, the security server checks authorization for the individual user and then, if appropriate, permits downloading of appropriate data from the database file server.

In another aspect of the invention, data for a molecular structure is loaded into the database by specifying the file pathnames for the various data files that contain the different types of data, including the different molecule views. Using a browser to view the data files permits various helper applications, called plug-ins, to smoothly and transparently accept the different file formats and provide views to the user. The various data files of the database are organized in accordance with the database design when they are loaded into the database and are managed by a relational database management program.

Other features and advantages of the present invention should be apparent from the following description of the preferred embodiment, which illustrates, by way of example, the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of a network system constructed in accordance with the present invention.

Figure 2 is a block diagram showing the primary functional components of the database server illustrated in Figure 1.

Figure 3 is a block diagram showing the primary hardware components of the database server and security server machines illustrated in Figure 1.

Figure 4 is a representation of a Login screen shown to a user at a client machine display of the network system illustrated in Figure 1.

Figure 5 is a representation of a selection screen shown to a user following proper login through the display illustrated in Figure 4.

Figure 6 is a representation of a selection screen shown to a user following selection of a database option from the display illustrated in Figure 5.

Figure 7A and Figure 7B show a representation of a protein families display screen showing an index of the protein families available for selection from the database server illustrated in Figure 1.

Figure 8 is a representation of a query submission screen shown to a user following selection of the query option from the display illustrated in Figure 6.

Figure 9A and Figure 9B show a protein database listing generated in response to a query submitted from the display illustrated in Figure 8.

Figure 10 is a representation of a query submission screen for a search based on a particular disease state.

Figure 11 is a representation of a protein information screen such as might be generated in response to a query or in response to selection of a protein from the protein name display illustrated in Figure 9.

Figure 12A, 12B, 12C, 12D, 12E, and 12F show protein structural data as stored in a PDB-format data file, comprising one of the data files stored in the database server illustrated in Figure 1.

Figure 13 is a representation of a protein Visualization Toolkit screen for a protein selected from the database server illustrated in Figure 1.

Figure 14 is a representation of a viewing screen displaying a 3-dimensional view of a protein structure selected from the database server illustrated in Figure 1.

Figure 15 is a representation of the functionality contained in the Measure menu showing a graphical display of the interatomic distance between two atoms of the protein molecule shown in Figure 14.

Figure 16 is a representation showing a graphical display of the angle between three atoms in the protein molecule shown in Figure 14.

Figure 17 is a representation showing a graphical display of the dihedral angle between four atoms in the protein molecule shown in Figure 14.

Figure 18 is a representation showing a graphical display of the atomic coordinates for an atom in the protein molecule shown in Figure 14.

Figure 19 is a representation of the Sequence Viewer window showing the amino acid sequence of the protein molecule shown in Figure 14.

Figure 20 is a representation of the Sequence Alignment window showing the alignment of the amino acid sequence of the protein molecule shown in Figure 14 with a template sequence.

Figure 21 is a representation of the Secondary Structure prediction window showing the amino acid sequence of the protein molecule shown in Figure 14 and the predicted secondary structural features of the protein.

Figure 22 is a representation of the Visualization menu selections for the protein molecule shown in Figure 14.

Figure 23 is a representation of the Secondary Structure Ribbon menu selections for the protein molecule shown in Figure 14.

Figure 24 is a representation of the Quality menu selections and shows a Ramachandran plot for the protein molecule shown in Figure 14.

Figure 25 is a representation of the Quality menu selections and shows a Balasubramanian plot for the protein molecule shown in Figure 14.

Figure 26 is a representation of the Quality menu selections and illustrates the Ellipsoid functionality for the protein molecule shown in Figure 14.

Figure 27 is a representation of the Surface Hydrophobicity menu selection for the protein molecule shown in Figure 14.

Figure 28 is a representation of the Strain Plot menu selection for the protein molecule shown in Figure 14.

Figure 29 is a representation of the Profile Analysis menu selection for the protein molecule shown in Figure 14.

Figure 30 is a representation of the Align functionality showing a list of proteins to select for superimposing on the protein molecule shown in Figure 14 and a window for displaying the superposition of multiple proteins.

Figure 31A and Figure 31B show a display screen for entering data about a protein into the database server illustrated in Figure 1.

Figure 32 is a flow diagram representation of operations performed when a user accesses the protein database of the system illustrated in Figure 1.

Figure 33 is a flow diagram representation of the operations performed during the security checking operation box of Figure 25.

Figure 34 is a representation of the database schema used by the relational database management system illustrated in Figure 1.

Figure 35 shows the object classes that produce the screen displays from the browser program at a user terminal illustrated in Figure 1.

Figure 36 is a block diagram representation of an alternative network system constructed in accordance with the present invention, using a distributed architecture and "Enterprise Java Beans" components.

Figure 37 is a block diagram representation of the "Enterprise Java Beans" components in the Figure 36 system.

Figure 38 is a representation of an Application screen shown to a user at a client machine display of the network system illustrated in Figure 36, with a Family Tree panel shown at the left side of the application window.

Figure 39 is a representation of the Application screen shown to a user at a client machine of the network system illustrated in Figure 36, illustrating a Search panel shown in the application window.

Figure 40 is a representation of the Application screen shown to a user at a client machine of the network system illustrated in Figure 36, illustrating a Data Mining panel shown at the left side of the application window with a Structure visualization panel at the right side of the application window.

Figure 41 is a representation of the Application screen shown to a user at a client machine of the network system illustrated in Figure 36, illustrating simultaneous Quality display features for two selected proteins in the Visualization panel of the application window.

Figure 42 is a representation of the Application screen shown to a user at a client machine of the network system illustrated in Figure 36, illustrating simultaneous Structure display features for two selected proteins in the Visualization panel of the application window.

Figure 43 is a representation of the Application screen of Figure 42, showing the drop down menu for selection of Display features.

Figure 44 is a representation of the Application screen of Figure 42, showing the drop down menu for selection of Options for atoms to be viewed.

Figure 45 is a representation of the Application screen shown to a user at a client machine of the network system illustrated in Figure 36, illustrating simultaneous

Sequence display features for two selected proteins in the Visualization panel of the application window.

Figure 46 is a representation of the database objects for the database design of the system illustrated in Figure 36.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The invention will be better understood with reference to the attached drawings, in which like reference numerals refer to like objects. For purposes of illustration, the system is described with respect to a database of protein structures. It should be understood that the system could also be adapted for use with databases of 3-D structures of other biological molecules or macromolecules, such as DNA, RNA or carbohydrates.

1. System Components

Figure 1 is a block diagram of a network system 100 constructed in accordance with the present invention. The system includes a database server 102 that communicates over a high speed communications line 104 with a security server 106. Multiple user client machines 108, 110, 112 are shown in communication with the security server 106 over high speed network communications lines 114, 116, 118, respectively, to gain access to data stored at the database server 102. The database server stores protein data in a relational protein database that can be searched by protein family and by a variety of protein characteristics, so that files relating to the same protein are stored in the same subdirectory. The files comprise a wide variety of protein data, and can include PDB-format files, text files, graphic image files, virtual reality files, video files, and audio files.

Each of the client machines 108, 110, 112 comprise a network terminal that permits a user to gain access and retrieve data using a resident browser on their respective client machine. The browsers provide a graphical user interface to access all of the collected database files and view the data without changing the interface.

In accordance with the invention, a user at one of the terminals 108, 110, 112 communicates over a public network connection 114, 116, 118 after receiving authorization from the security file server 106, which controls access to the collected files. Access to the security server and collected files at the database server 102 is provided through the graphical user interface of the conventional browser program that is widely available, such as the Communicator or Navigator browser product from Netscape Communications Corp. or the Internet Explorer browser product from Microsoft Corporation. In this way, the invention provides a convenient means of searching among protein families for proteins with characteristics of interest. The invention also permits data searching, file viewing, and investigation of multiple views of proteins from within a single viewing program.

Each of the network terminals 108, 110, 112 comprise a computing platform that typically includes a central processor unit (CPU), such as the Pentium-class of microprocessor chips manufactured by Intel Corporation or the CPU microprocessors manufactured by Silicon Graphics, Inc. The terminals also include a video display unit, such as a video monitor or other display screen, and a network interface. In the preferred embodiment, each network terminal also includes another means of accessing data on storage media, such as a floppy disk drive or a CD-ROM or DVD-ROM drive. Because

each network terminal 108, 110, 112 communicates with the security server 106 using the same type of graphical user interface, access to the security server does not depend on the operating system of each terminal being the same. For example, each of the illustrated terminals 108, 110, 112 may use a different operating system. Thus, the first client machine 108 may function using the "Windows 95" operating system program from Microsoft Corporation. The second client machine 110 may function using the Unix operating system, and the third client machine 112 may operate using the Apple MacIntosh operating system. Access to the security server will be transparent to the users at these client machines 108, 110, 112 as long as the browser at each terminal is operable with the respective client operating system.

As noted above, the client network terminals 108, 110, 112 communicate through a web browser interface to the security server 106 and then gain access to the database at the database server 102. Figure 2 is a block diagram of the database server 102. The database server is a computing machine that supports a relational database system 202, such as the systems manufactured by Oracle Development Corporation, which accesses the collection of data files 204 comprising the protein database of the system 100. That is, the relational database system 202 provides access, security control, and search and retrieval services for the database files 204 that are stored at the database server 102. The database files may include a wide variety of file types, including alphanumeric text files, graphic image files, video files, and audio files.

The database server 102 also includes a web application server 206, which interfaces with the relational data base system 202 to retrieve and, if necessary, process

the data files 204. The web application server may include programs known as cartridges. Those skilled in the art will be familiar with cartridges and with the tasks they can perform, and will understand that cartridges specific to particular database systems can be readily developed by users of the database system 202. As described further below, for example, cartridges are used to control access to the database, so that data subscribers of different authorizations are provided with different access levels to the files, or will be granted access with different exceptions. The database server 102 also includes a controller 208, which provides the operating system at the server. The operating system in the preferred embodiment supports various program types, including those written in the "Java" programming language developed by Sun Microsystems, Inc of Palo Alto, California, USA.

Figure 3 is a block diagram of the primary hardware components of the computers that comprise both the database server 102 and the security server 106. It should be understood that the functions of the database server and the security server can be implemented on the same computer, if desired. In the preferred embodiment, the functions are performed at separate computers to facilitate database maintenance at the database server and increase the rate at which new data can be uploaded and made available at the database. In addition, the client network terminals 108, 110, 112 may each have a construction similar to that shown in Figure 3. As noted above, each client machine 108, 110, 112 may function with a different operating system, as may the servers 102, 106, the only requirement being that they can communicate with each other over the network connections 104, 114, 116, 118. The exemplary computer 300

includes a CPU 302 that provides substantial computing power to successfully handle a large volume of user traffic and to manage transfer of large amounts of data. For example, when a user requests the database files for a single protein, the amount of data transferred from the database server 102 can comprise more than 2 megabytes (MB) of data. The computer 300 also includes memory 304, which typically includes 64 MB or more of fast random access memory (RAM). The computer will also include a network interface 306 to permit high speed communications with a network 308, such as provided by direct connection with T1 data lines, a frame relay connection, or other high bandwidth digital data line. The computer also will include an external storage device interface 310 to permit transfer of data, including programming, from external storage media 312, such as floppy disks, magnetic tape, CD-ROM, and DVD-ROM storage media. Internal data storage, of persistent data too large to fit within the memory 304, may be stored in disk storage 314. Presently, readily available disk storage provides space of six to eight gigabytes (GB).

A user will initiate action and input data from a combination of input devices 320, such as a keyboard and display mouse, and will view data on a display 322 such as a video monitor. Typically, the database server 102 and security server 106 will be more powerful machines than the network terminals 108, 110, 112, because the servers must potentially handle a large amount of communications traffic. The network terminals must be able to communicate with the security server and receive large amounts of data, but the required speed of such data transfer is to be determined by the user preference. It has been found, however, that a network connection of greater than 56K/sec data

transfer is most desired, and slower terminal-to-security server communications will not be satisfactory to most users.

2. Using the System to Retrieve Data

The first step in gaining access to the stored database and the data files contained therein is to obtain access to the database server 102. To gain access to the server, a user must first receive access through the security server 106 by standard conventional communication techniques. For example, the Internet protocol (IP) address of the security server must be known, so that a user 108, 110, 112 can communicate with the server 106 through the user's browser program. Upon communicating with the security server 106, the user is presented with the first login screen at the browser interface. Figure 4 is a representation of a login screen 402 that is shown to a user 108, 110, 112 (Figure 1) at a client machine display unit, viewing through the user's conventional browser window. The login screen identifies the database to be accessed, and provides a user with a "login" request button 404 to be selected (clicked) using the client machine display mouse, keyboard, or other window designation device. No program response will occur until the login button is selected.

After a user clicks on the login button 404, the security server will provide the user's browser with a display screen that shows a list of database selection features, or utilities, from which one must be chosen. Figure 5 is a representation of the information box screen 502 that is shown to a user following selection of the login button in the screen display illustrated in Figure 4. In the Figure 5 representation, two of the illustrated database selections are shown for variety, but do not form any part of the invention; these

are the selections for "SVdBase" 504 and the "CombiLib" 506. Other permitted database selections may be included. The other two database selections shown in Figure 5, for "SBdBase" 508 and for "SBdBasePlus" 510, are used to select access to the unique relational database of protein structural data provided in accordance with the present invention. The two choices 508, 510 select different levels of access to the database files. The different levels of access may be, for example, to different protein families, so that certain protein families may be made available only at a higher fee. The different levels of access also may restrict access to files within a given protein family, so that certain types of data may be made available only at a higher fee.

After a user selects one of the database access levels in Figure 5, the security server will present the user with a database selection screen. Figure 6 is a representation of the database selection screen 602 shown to a user following selection of the database option from the display illustrated in Figure 5, and shows that a user can select between an index view display button 604 and a query display button 606. If the index view button 604 is selected by a user, then the security server will return a list of protein families that are available for viewing. In the preferred embodiment, all protein families are shown in the index view, according to protein family name, even if the user does not have authorization to view them. In this way, users will be kept informed of the full database that is potentially available, including recently added data for which they may not have access. Thus, users may decide to upgrade their level of access upon viewing the complete list and determining the available proteins to which they do not currently have access rights. Figure 7A and Figure 7B comprise a display screen list 702 that

shows an exemplary index listing of the available proteins displayed by selecting the index button 604. If the user selects the query display button 606 from the Figure 6 display, then the security server will return a query screen so the user can designate a query on the database.

Figure 8 is a representation of a database query display screen 802 shown to a user following selection of query submission from the display screen illustrated in Figure 6. The query screen 802 shows a display area 804 that lists query fields over which searching can be performed. In the preferred embodiment, these fields include a database identification (ID) number, protein name, species, gene, conventional "SwissProt" accession number, disease state, protein function, protein family, and full-text search string. After the user has determined the search query, the user selects the "Submit Query" display button 806 and the user's browser sends the query information to the security server.

Figure 9A and Figure 9B illustrate a query results display screen 902 that is shown to a user following selection of the query option from the display illustrated in Figure 8, where a user has selected the "Protein Name" field for search. Thus, the Figure 9A and 9B display shows a list of protein names, corresponding to the names of proteins available in the database at the database server 102 (Figure 1). That is, the left-most column 904 in Figure 9A, 9B lists the protein name, the next column 906 shows the corresponding database ID number, the next column 908 contains an indication (where applicable) of whether PDB data is available, and the last column 910 contains an

indication (where applicable) of whether homology data is available for the named protein.

Figure 10 is a representation of the query display screen first illustrated in Figure 8, except that the Figure 10 display screen 1002 shows that "cancer" has been inserted into the "Disease" state field 1004 so that a query is executed that will search for proteins related to cancer. More particularly, when the "Submit Query" display button 1006 is selected in Figure 10, the security server will receive the "cancer" disease query and will cause a search to be executed over the database files by the database server 202 (Figure 2). The search results will be returned by the security server to the user's browser.

Figure 11 is a representation of a protein information screen 1102, such as might be generated in response to a query or in response to selection of a protein from the protein name display illustrated in Figure 9. The Figure 11 screen shows that, in the preferred embodiment, the protein information includes the database ID number 1104, the complete protein name 1106, the species 1108 for the protein data file, the gene type 1110, predetermined database keywords 1112 for the protein, disease information 1114, function information 1116, protein family class 1118, the SwissProt accession number 1120, and the EC number 1122. These data contained in these fields will be familiar or readily understood to those skilled in the art. A row of display selection buttons below the information fields of Figure 11 show that a user may select to retrieve the corresponding data file for the protein from the PDB-format database 1130, or may select the SwissProt data 1132 for the protein, or may select "GenBank" data 1134, or may select "PIR" data 1136 for the protein. Of these alternatives, it should be noted that only

the PDB-format database contains multiple data types, other than protein sequence data. That is, the SwissProt, GenBank, and PIR databases contain only sequence data and not molecular structure data. The user may decide to begin another query by selecting the "New Query" display button 1138.

3. The Protein Visual Display Tools

As noted above, protein structure data is alphanumeric data that can be searched using text strings, but is not meaningful to researchers without the aid of viewing programs. Figure 12A, 12B, 12C, 12D, 12E, 12F (collectively referred to as "Figure 12") is a representation of protein structure data as stored in a PDB-format data file, comprising one of the data files stored in the database server illustrated in Figure 2. The standard PDB format of the protein data shown in Figure 12 will be familiar to those skilled in the art. It should be apparent that such data is tedious to review and is not especially revealing of protein structures and characteristics. The database system of the present invention organizes such data into a coherent database representation and permits a visual interface to such data through a conventional Internet web browser interface.

Figure 13 is a representation of a protein "Visualization Toolkit" display screen 1302 for a protein selected from the database server 202 (Figure 2) using the interface of the present invention. In response to the user selecting a protein name from the "database" button of Figure 11, the security server causes the visualization toolkit interface program to be launched and displayed within the user's browser, and returns the Figure 13 display page as the opening page of the visualization toolkit interface program. Thus, a user will always be provided with the display of Figure 11 after selecting a

protein through either the protein name displays or from executing a query, so that selection of the visualization toolkit may be made for the desired protein.

Figure 14 is a representation of a viewing screen displaying a 3-dimensional view of a protein structure selected from the database server illustrated in Figure 1. When a structure is read into the viewer, the atoms are displayed and can be colored according to atom type. The structure can be manipulated within the 3-D graphics window using the computer mouse to control its movement. Across the top of the screen are six pulldown menu selections under which are commands for structure visualization and analysis. In some cases, commands can only be executed if precalculated information is stored in data files associated with a particular protein in the database. If such information is not available for a given protein, the command that requires the information will be blanked out and will not be accessible for that protein. Shown at the right side of the screen is the atom identifier information for any atom selected in the structure (e.g. K 224, CA), as well as commands for clearing any screen annotations, downloading coordinates for additional structures, for accessing the database index and for submitting a new database query.

Figure 15 is a representation of the functionality contained in the Measure menu. The figure shows a graphical display of the interatomic distance between two atoms in the protein, which is created by selecting the Distance command and picking the two atoms using the cursor and mouse. The Angle (Figure 16) and Dihedral (Figure 17) commands calculate and display angle values when the user picks either three or four

consecutive or non-consecutive atoms in the structure. The Coordinate (Figure 18) command displays the atomic coordinates for a selected atom.

The View pulldown menu includes functionality for displaying selective atoms in the protein, for example, all atoms or only main chain or alpha carbon atoms.

The Sequence menu includes commands for amino acid sequence analysis. The Sequence command brings up a separate window in which the protein sequence is displayed according to single letter amino acid codes and colored according to residue type. Figure 19 shows a representation of the Sequence Viewer window showing the amino acid sequence of the protein molecule shown in Figure 14. The sequence and structure windows are interactive in that placement of the cursor on an amino acid code in the sequence highlights the position of that amino acid in the structure. The sequence window is independent of the main molecular structure window in that it can be moved, resized and closed as a separate frame. These capabilities are due to functionality in the "Java" language and are applicable among all windows present in the system.

The Alignment functionality (Figure 20) shows the alignment of the protein molecule with one or more template sequences. The sequence alignment is calculated using any one of a number of sequence alignment algorithms known to those of skill in the art, for example, programs such as MSA (Carrillo and Lipman, "The multiple sequence alignment problem in Biology", SIAM J. Appl. Math. 48, 1073-1082 (1988); Altschul and Lipman, "Trees, stars and multiple biological sequence alignment", SIAM J. Appl. Math., 49, 197-209 (1989); Altschul, "Gap costs for multiple sequence alignment", J. Theor. Biol., 138, 297-309 (1989); Altschul et al., "Weights for data

related by a tree", J. Molec. Biol., 207, 647-653 (1989); Altschul, "Leaf pairs and tree dissections", SIAM J. Discrete Math., 2, 293-299 (1989); Lipman et al., "A tool for multiple sequence alignment", Proc. Natl. Acad. Sci. USA, 86, 4412-4415 (1989)) or ClustalW (Higgins et al., CABIOS, 8, 189-191 (1991)), which are available in the public domain, can be used.

The Secondary Structure command (Figure 21) opens an additional window in which is displayed information about the predicted secondary structure of the protein, for example, whether a particular residue is involved in a helix, coil or sheet. This information is precalculated and stored in the database, for example, by using a publicly available algorithm for calculating secondary structure, such as SSPAL (Salamov and Solovyev, "Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments", J. Mol. Biol. 247, 11-15 (1995)) or can be calculated interactively.

Figure 22 is a representation of the Visualization menu selections for the protein molecule shown in Figure 14. This menu contains functionality for visualizing different structural attributes of the displayed molecule. The Hydrophobicity command interactively colors the residues of the protein according to a color scheme based on a predetermined hydrophobicity scale for the individual amino acid residues. This allows identification of hydrophobic and hydrophilic regions in the protein structure. The Active Sites, Glycosylation, Phosphorylation and Natural Variants commands display these sites in a protein molecule, if known, in accordance with data which has been previously determined and stored in association with the particular protein.

The Secondary Structure Ribbon command (Figure 23) opens a window in which is displayed a solid ribbon along the protein backbone highlighting the secondary structural attributes of the protein, for example, helices, sheets or coils. The ribbon is precalculated, such as by using the publicly available Ribbons program (Carson and Bugg, "Algorithm for ribbon models of proteins", J. Mol. Graphics, 4, 121-122 (1986); Carson, "Ribbon models of macromolecules", J. Mol. Graphics, 5, 103-106 (1987); Carson, "Ribbons 2.0", J. Appl. Cryst., 24, 958-961 (1991); Carson, "Ribbons", Methods in Enzymology, R.M. Sweet and C.W. Carter, eds, Academic Press, 277, 493-505 (1997)) and is stored in the database in a data file associated with a given protein. The solid surface is displayed in the graphics window, for example, by plugging in a utility such as the SGI Cosmo Player to Netscape. Also precalculated and stored along with the protein are electrostatic and dynamic surfaces, which are called up and displayed as solid surfaces in a separate window using the Electrostatic Surface and Dynamic Surface commands, respectively. The electrostatic surface is precalculated using an available program, such as GRASP (Nicholls et al., PROTEINS, Structure, Function and Genetics, Vol. 11, No. 4, pg 281ff (1991)), for calculating electrostatic potential. The dynamic surface is calculated from molecular dynamics data, which is calculated by using any number of molecular dynamics software packages. Such packages are well known to those of skill in the art. The Dynamic Surface command color codes the residues in the protein according to movement or flexibility during molecular dynamics simulation. For example, "hot" residues move the most and are colored red; residues that move over

average trajectories are colored green; and the "cool" residues that exhibit the least movement are colored blue.

Figures 24-29 are representations of the Quality menu selections. The commands in the Quality menu are used to evaluate the quality of a model structure by validating that the structural and energetic characteristics of the molecule are within a reasonable expected range of values. Figure 24 shows a Ramachandran plot for the protein molecule shown in Figure 14. The plot is calculated interactively when the command is executed and displays the phi and psi angle values for the protein structure in a separate window.

Figure 25 is a representation of the Quality menu selections and shows a Balasubramanian plot for the protein molecule shown in Figure 14. The Balasubramanian plot is also calculated interactively for the displayed protein and indicates a directional arrow between the phi and psi angle values for each residue in the protein. Each residue is color coded according to secondary structure based on the dihedral angle values, for example, residues involved in helices are colored red and those in beta sheets are colored blue.

Figure 26 is a representation of the Quality menu selections and illustrates the Ellipsoid functionality for the protein molecule shown in Figure 14. The ellipsoid displays the total extent or size of the protein in the x-, y- and z-directions.

Figure 27 is a representation of the Surrounding Hydrophobicity for the protein molecule shown in Figure 14 as calculated using the Surface Hydrophobicity command. The hydrophobic packing for the protein is precalculated based on how far a residue is from the protein surface (Ponnuswamy and Prabhakaran, "Properties of nucleation sites

in globular proteins", Biochem. and Biophys. Res. Comm., 97, 1582-1590 (1980); Ponnuswamy et al., "Hydrophobic packing and spatial arrangement of amino acid residues in globular proteins", Biochim. Biophys. Acta, 623, 301-316 (1980)), and the result is plotted in a separate window. The interior residues are displayed in the center of the plot in the area of highest hydrophobicity and the surface residues are shown at the edge of the plot, indicating areas of lowest hydrophobicity. This plot can give information on nucleation sites and can be compared to crystal structure data.

The Strain Plot and Local Strain Energy commands display plots of internal strain energy in the molecule. The strain plot (Figure 28) is displayed in a separate window as a graph of strain energy per residue. The local strain energy is displayed in a separate window as a solid ribbon along the protein backbone which is colored according to strain energy (for example, highly strained residues are colored red and unstrained residues are colored blue). Strain energies are precalculated by using external programs, such as ICM, and stored in the database in association with the protein structure. The Profile Analysis command (Figure 29) displays a plot of packing factor per residue, which is precalculated using a publicly available program such as WHAT IF (Vriend, "WHAT IF: a molecular modeling and drug design program", J. Mol. Graph. 8, 52 (1990)) and stored in a data file in the database. The profile analysis is displayed in a separate window as a graph of packing factor per residue.

Figure 30 is a representation of the Align functionality. The Align command aligns proteins within subfamilies. A list of proteins within the subfamily is displayed in a viewer window and can be selected for superimposing on the protein molecule

shown in the viewer window. A separate window displays the superposition of the selected proteins.

4. Entering Protein Data Into the System Database

Figure 31 is a representation of a display screen for entering data about a protein into the database server illustrated in Figure 1. As noted above, the database design of the system illustrated in Figure 1 collects files of different types, such as text, graphic, video, virtual reality, and audio. The database design further collects data files and places them in a relational organization, so that files that are related to the same protein family are readily accessible through a search routine. In accordance with known relational database management programs, such searching may be carried out with specialized search languages, such as Structured Query Language (SQL). In the preferred embodiment, data is entered into the database by supplying pathnames to data files that are loaded into the data storage of the database server 102 (Figure 1). The template for entering such pathname information is supplied by the entry display screen 3102 of Figure 31. The display screen includes fields that accept input that is manually entered by an authorized user, as well as data file names to identify protein data files that will become interrelated by the database management system.

As indicated in Figure 31, the first data field 3104 is for the protein family name, followed by the protein name 3106. This information is manually entered by a user who gains access through the security server or other special access means at the database server. When the information is received by the database management program at the database server, it is automatically stored according to the database design. The next

information that is manually entered is for the species name 3108. Following the species name, the user who is entering the database information must supply data file names. The file names will comprise a data file pathname to a data file stored at the database server.

The first data file pathname 3110 is to an annotation file, which is a text (alphanumeric) file that contains predetermined supplementary protein data, such as disease information, ID numbers, and the like. Those skilled in the art will understand that a text file has a standardized file extension of ".txt" in the filename. The annotation file provides a means of collecting the supplementary protein data in one convenient file, and may be created by a user manually entering the necessary data, or may be the result of processing other files or data, such as the protein PDB data file.

The next file pathname to be entered into the database is for an alignment text file 3112 that contains atomic alignment data. The next pathname is to a secondary structure file 3114, which is another text file. A ribbon setup file pathname 3116 and a ribbon data file pathname 3118 are next entered. These data files comprise virtual reality files, which have a standardized file extension of ".wrl" and which are viewed in a web browser program using any one of a variety of readily available plug-in applications. Those skilled in the art will appreciate that a standard file extension such as ".wrl" is recognized by operating systems and browsers to automatically trigger launch of a program that can view the data. For example, one ".wrl" viewer program that integrates with conventional Internet browsers such as Netscape Navigator is the "Cosmo Player" by Platinum Technology, Inc. for playing virtual reality files that are coded in accordance with the

VRML 2.0 open standard for Internet programming. The next entered file pathname is to a natural variant coordinate file 3120, which is a text file.

The next data pathname to be entered is for an electrostatic surface file 3122, which is a virtual reality ".wrl" file. Those skilled in the art will understand the type of molecule view that corresponds to an electrostatic surface. The next file is a text file that contains surface hydrophobicity data 3124. Next is a profile analysis data file 3126, which contains atomic coordinate data in the same format as a PDB file, but which is created with homology model techniques. Next is a protein ellipsoid data file 3128, a text file. An accessibility data file 3130 is next, comprising a text file.

A local strain data file 3132 is the next file pathname, and is a text file. Thereafter, the next two files relate to local strain data. The first is a local strain ribbon data ".wrl" virtual reality file 3134, and the second is a local strain ribbon line data ".wrl" virtual reality file 3136. The next pathname is to another ".wrl" file, a dynamic surface file 3138. The dynamic surface file provides a view of the molecular outer surface of the protein.

The next items for database input relate to the initial model 3140. The first entry is the pathname of a text file comprising an initial coordinates file 3142. The next three items shown in Figure 31 are optional, in that the information they specify can be extracted in real-time from other data files by the visualization toolkit interface program. If the visualization toolkit interface program has the ability to extract such data in real-time, then the optional file pathnames shown in Figure 31 need not be supplied. These optional files comprise a Ramachandran plot 3144, bond length graph 3146, and bond

angle graph 3148. Those skilled in the art will appreciate that these three types of views can be easily generated in real-time, without any need for pre-processing. The last data entry relating to the initial model is for the depositor name 3150. This field identifies the person who is entering the data, and is useful for identifying persons who should be able to answer questions about the particular data files they entered.

It is possible to review the atomic coordinate data for a protein and, after careful consideration by expert review, it is possible to improve the molecular representation by modifying the coordinate data. Therefore, the preferred embodiment of the system accepts multiple levels of data refinement, as indicated in Figure 31 by the "Refinement Level 1" 3160 and "Refinement Level 2" 3162 entry fields. Thus, Level 1 includes a field for the pathname of the Level 1 coordinates file 3164, optional fields for Level 1 refinement Ramachandran plot pathname 3166, Level 1 refinement bond length graph pathname 3168, and Level 1 refinement bond angle graph pathname 3170, as well as the mandatory field for Level 1 refinement depositor name 3172. Similarly, the Level 2 fields 3162 include a field for the pathname of the Level 2 coordinates file 3174, optional fields for Level 2 refinement Ramachandran plot pathname 3176, Level 2 refinement bond length graph pathname 3178, and Level 2 refinement bond angle graph pathname 3180, as well as the mandatory field for Level 2 refinement depositor name 3182.

Finally, at the bottom of the data entry display of Figure 31, the entered data can be submitted by clicking the "Submit" display button 3190. The entered information can be cleared by selecting the "Clear" button 3192.

5. System Operation

Figure 32 is a flow diagram representation of operations performed when a user accesses the protein database of the system illustrated in Figure 1. The first operation is for the security server to perform a security check to ensure that the user has authorization to proceed with access. This operation is represented by the Figure 32 flow diagram box numbered 3202. As noted above, different users may be granted different levels of access. This is described further below. After a user has been granted access, the next operation is for the user to select a database for viewing. The request must be authorized by the security server. This operation is represented by the flow diagram box numbered 3204, and the screen display being viewed by the user corresponds to Figure 5.

Next, the user selects either a search query or an index list of protein names available in the database. At this point, represented by the flow diagram box numbered 3206, the user is viewing a screen display like that of Figure 6. Selection of the index list by a user will cause the user's network terminal to initiate a request to the database server to execute a corresponding database server cartridge program to create and display the index list of available protein names. In the preferred embodiment, the entire list of available SBdBase protein names is shown to all users, regardless of authorization level. The protein names available to the user may be indicated by special formatting, such as boldface or special characters. In this way, the user can be appraised of new database entries and possibly interesting protein names that are unavailable. This can assist in

migrating users from a lower level of authorization to a higher level of authorization for viewing the proprietary database.

Next, the user selects a protein name for viewing, either from an index list or from the results of a query search. At this step, once a protein name has been selected, the user will be shown a display like that of Figure 11. Selecting a protein name causes the user's network terminal to initiate a request to an appropriate web application server to create the display page corresponding to Figure 11. In accordance with the web application server programming, for example of the kind available from Oracle Development Corporation, the proper display page is automatically created. This operation is represented by the Figure 32 flow diagram box numbered 3208. The user may select from multiple database choices. As illustrated in Figure 11, these choices may include the "SwissProt" database, "GenBank" database, or PIR database, all of which contain only text data. Alternatively, the choice may be for the "SDdBase", comprising the multi-format protein relational database constructed in accordance with the present invention. The selection of the "SBdBase" is indicated by the flow diagram box numbered 3210.

With selection of the "SBdBase" button, the user's network terminal sends a request to the security server, which recognizes the "SBdBase" selection and initiates execution of appropriate "Java" programming language scripts. Such scripts are executed quickly and without database server involvement, and so provide convenient and more efficient operation of the system. This operation is represented by the flow diagram box numbered 3212. The Java scripts cause the relevant database pathnames to be provided

from the security server to Java language programming routines at the database server controller. The pathnames may include, for example, the pathnames corresponding to the multi-format data files entered into the database, such as illustrated in Figure 31. The pathname sending operation is represented by the Figure 32 flow diagram box numbered 3214.

When the database controller receives the database pathnames, the controller causes the associated database files to be sent from the database server through the security server and on to the client network terminal, thereby causing the database files to be downloaded to the network terminal. This operation is represented by the flow diagram box numbered 3216. After the files have been downloaded to the network terminal, the user at the terminal can invoke the "Visualization Toolkit" and view all associated displays (comprising the views illustrated in Figures 14 through 30). Such viewing will be accomplished through the user's web browser program, as described above, and comprise entirely local operations. Thus, network traffic will not be involved. These local operations are represented by the Figure 32 flow diagram box numbered 3218. Thus, viewing operations can be continued locally, as indicated by the "Continue" box 3220 of Figure 32.

Figure 33 is a more detailed flow diagram representation of the operations performed during the security checking operation box 3202 of Figure 32. In Figure 33, the first operating step is represented by the sending of a request for access from the user's network terminal to the security server. This operation is represented by the flow diagram box numbered 3302. In the preferred embodiment, the security server checks

the Internet protocol (IP) address of the sending user against an authorization list and grants conditional privileges accordingly. This operation is represented by the flow diagram box numbered 3304.

Next, the security server checks the received user login information for verification. This may comprise, for example, checking user name and user password information received from the network terminal. This security operation is represented by the flow diagram box numbered 3306. If all security checking is approved, then the security server grants access and returns appropriate display screen information to the user, as represented by the flow diagram box numbered 3308. The system operation then continues as described in Figure 32.

As noted above, a relational database management system (RDBMS) provides access, security control, and search and retrieval services for the database files stored at the database server. Figure 34 is a representation of the database schema used by the relational database management system illustrated in Figure 1. The database schema defines the tables that will be used by the RDBMS in performing the services described above. Each protein in the database will have data in the RDBMS tables corresponding to the fields described in Figure 34. Thus, the tables in Figure 34 determine the fields over which the RDBMS can perform the search and retrieval services and determine the level of control exercised over the other services provided by the RDBMS. In the preferred embodiment, the tables are implemented by the RDBMS produced by Oracle Development Corporation. Those skilled in the art will understand that RDBMS products from other vendors are equally suitable.

The primary table in Figure 34, called MY_CORE, includes fields from which the menu display of Figure 5 is created. The MY_CORE fields include a field for a specific protein identification number, listed as SP_ID, unique to the database implementation, for internal identification. Other fields are for accession number (ACC_NUM), pdb-file indicator (PDB_FLAG), database flag for the additional views provided by the invention (SBDBASE_FLAG), special code (SPEC_CODE), a description field (DESCRIPTION), disease notes (DISEASE_NOTES), function notes (FUNCTION_NOTES), sequence data (MY_SEQUENCE), and keywords for searching (KEYWORDS). The MY_CORE table also includes fields for gene data (GENES), enzyme data (ENZYME), pdb data (PDB_DATA), and notes to indicate similarities to other proteins (SIMILARITY_NOTES).

Another table for the RDBMS is called FAMILY, which contains a single field for the family name of the protein to which the Figure 34 tables correspond. Another table in Figure 34 is called SUBFAMILY, which contains the family name and also any subfamily name for the protein. A PROTEINS table contains fields for protein family name, subfamily name, protein name, and protein identification number (SP_ID). Thus, a system user can search the database, using the RDBMS, and search for protein name, protein family, protein subfamily, and protein identification number.

The RDBMS tables also include a USER_ACCESS table, with which the RDBMS controls access to the database depending on the individual user. That is, for each protein entry in the database, the USER_ACCESS table indicates whether a particular user has been granted access. As noted above, a user can be granted viewing

access to the database, protein by protein. Thus, the USER_ACCESS table has fields for user name (USER_NAME), protein identification (SP_ID), and a conventional protein identification number.

The CUSTOMERS table contains information that is used by the RDBMS to control database access according to customer accounts. The fields for CUSTOMERS include USER_NAME, START_DATE, and END_DATE. Thus, customer access is controlled by time period, to reflect whether a customer account is current or past due for payments.

If the information in the Figure 34 tables are used for search and retrieval, then the RDBMS uses the information in the fields to perform searching. Once proteins or other database entries are identified as satisfying the search request, the RDBMS retries the data files from the database server as described above and the data files are downloaded to the user.

The browser display technique described above is advantageous in that a wide variety of data formats can be accessed and displayed from within the same viewing program, independently of the operating system being used, as long as the browser has been configured with the appropriate helper applications. Those skilled in the art will appreciate that conventional browser programs are developed with object-oriented programming techniques, and that such browsers can be made to display the protein database information in an appropriate manner through proper interface with the browser programs. In particular, the Visualization Toolkit display shown in Figure is an application that executes from within the browser and provides the special menus and

window views shown above in the drawing figures. Figure 35 shows the classes that interface with the browser, in a manner specified by the browser vendor.

Figure 35 indicates object classes for the browser product from Netscape Communications Corporation, but other suitable browsers may be used and will occur to those skilled in the art. The classes include an Ellipsoid class that generates the ellipsoid view from the View menu. The class HydrophobicityPlot generates the hydrophobicity plot view. The class "netscape.application.InternalWindow" generates a window within the browser, according to the Netscape specification for the class netscape.application.Window. The subclasses for the internal window include windows for the Baluchandran plot (class sbiBaluCanvas), the hydrophobicity internal window (class sbiHydrophobicity), the List window (class List), the Profile window (class sbiProfile), the Protein Viewer window (class sbiProtViewer), the Ramachandran window (class sbiRama), the Sequence Viewer (class sbiSeqViewer), and the Strain window (class sbiStrain).

Other classes are for display windows or to perform calculations on-the-fly to generate data that is displayed in windows. Such classes include generating the Profile Plot (class ProfilePlot to generate data for the sbiProfile class), generating the Strain Plot window (class StrainPlot to generate strain data), a graph test function (class graphTest), generating the hydrophobicity graph (class hydrophobicityGraph to generate data), and generating the Profile data (class profileGraph). Other classes provide the Alignment window (class sbiAlignViewer), generating the Balucandran data (class sbiBalu), producing the Baluchandran window buttons (class sbiBaluButtons), providing the

Ellipsoid window (class sbiEllipsoid), and providing the browser frame (class sbiGui). Another class provides the pdb "Canvas" viewer window (class sbiPdbCanvas). Those skilled in the art will recognize that "Canvas" is a particular viewing program for data. Another class provides the pdb-data viewer (class sbiPdbViewer), and another class provides the window frame slider control (class sbiSlider). Finally, other classes provide the Active Sites display (class sbiActiveSites) and provide a converter application (class sbiConverter).

6. Alternative Embodiment with a Distributed Network Architecture

Figure 36 shows the configuration of an alternative network system 3600 constructed in accordance with the present invention, using a distributed network architecture and "Enterprise Java Beans" (EJB) components in a multi-tiered configuration. The distributed architecture may be characterized as comprising multiple programming tiers. This characterization of an EJB implementation will be familiar to those skilled in the art. See, for example, Client/Server Programming with JAVA and CORBA (2nd edition), by Robert Orfali and Dan Harkey, pp. 33-50.

The distributed system 3600 permits multiple client machines 3602, designated Web Client 1, Web Client 2, ..., Web Client n, in a first tier to communicate over a shared network 3604, such as the Internet, with a second tier comprising an Authorization/Security access server 3606 that controls access by the clients 3602 to a bioinformatics database. The access server 3606 can comprise one or more programs and machines that perform the duties of the security server 106 and database server 102 of the embodiment illustrated in Figure 1.

The client machines 3602 execute one or more user interface applets to interface with the Authorization server 3606, which communicates with multiple "Enterprise Java Beans" (EJB) components 3608 that provide the functionality needed to generate the display panels and features illustrated in Figures 37 through 45 for the second embodiment, as described further below. The Authorization server and EJB components form the second tier of networking and communicate with a third tier of the system, a Bioinformatics Database Management System (DBMS) server 3610. The Bioinformatics DBMS server manages the collection of protein data stored in a database and provides the protein data in response to requests and queries from the users 3602.

Those skilled in the art will understand how the EJB components 3608 can be implemented using a distributed object model of the database 3610. For example, the database can be structured to communicate with the clients according to the object communication standard called Common Object Request Broker Architecture (CORBA), which is specified by the industry consortium called Object Management Group (OMG). The CORBA standard will be familiar to those skilled in the art of database design.

Figure 37 is a block diagram representation of the classes into which the EJB components 3608 are organized. These components provide the functionality for the GUI of the second embodiment, as described further below. The functionality of these components generally duplicates the functionality of the corresponding classes described above for the first embodiment of Figure 13 through Figure 35. Figure 37 shows that the classes include a Java class for an Alignment database, to provide alignment views of

proteins when requested by the client user, and also include a Java class for an Atomic database, to provide atom-specific data.

The EJB classes 3608 also include classes for protein Chains, Domain, Family, Protein, Residue, Secondary Structure, Subfamilies, Subfamily Proteins, Users, VAST, and VRML The VRML class provides support for the "Virtual Reality" display system which, in the preferred embodiment, is achieved through the "Cosmo" VRML player interface developed by Silicon Graphics, Inc. of Mountain View, California USA. A Deployment EJB class handles communications tasks between user clients and the application server for activation of classes. The VAST component performs processing to check the database 3610 for protein similarity, such as for data mining functions. This information is used in generating the displays for local strain energy, secondary structures dynamic surface, and hydrostatic surface.

In the preferred embodiment, the VAST component provides an interface to a portion of the database 3610 (see Figure 36) that is constructed using the "Vector Alignment Search Tool" (VAST) search program that is publicly available from the National Institutes of Health (NIH) in Bethesda, Maryland, USA. In the preferred embodiment illustrated in Figure 36, the database includes VAST output for the database proteins. Thus, the VAST program is executed on the database proteins to create a database of protein similarity information, in accordance with the VAST standard. In this way, the VAST EJB component provides an interface to the VAST output data, such that a similarity search request from a system user can be provided by appropriately scanning the VAST database, rather than attempting to execute a comparison operation in real

time. This improves the response time of the system. Those skilled in the art will be familiar with the VAST search methodology, details of which are available from the National Center for Biotechnology Information (NCBI) division of the National Library of Medicine (NLM) at the NIH (available on-line through the World Wide Web at www.ncbi.nlm.nih.gov).

Figure 38 is a representation of an Application screen 3800 shown to a user at a client machine display of the network system illustrated in Figure 36. The Application screen is shown in a display window at a client machine, such as in a Java applet window. Those skilled in the art will understand the automatic launch of Java applets from a web browser. In the preferred embodiment of the distributed architecture system, the applet window includes a menu bar 3802 along the top of the applet window, a Protein Selection display panel 3804 along the left side of the applet display, with a sub-panel 3806 for showing proteins selected by the user and available for viewing, and a Visualization panel 3808 at the right side of the applet display.

In Figure 38, the Protein Selection panel 3804 shows a Family Tree display that lists the protein families that may be selected for investigation by the system user. As described further below, other display tabs in the Protein Selection panel indicate that a Search panel and a Data Mining panel may be called up, in addition to the Family Tree panel shown. In this regard, it should be noted that this embodiment of the invention permits especially rapid review and easy selection of protein families without need for the multiple window displays associated with the first embodiment described above in conjunction with Figures 5 through 9. A user can select protein families by positioning

computer the display cursor on a desired protein family in the Family Tree display and then using a display mouse button to "double-click" and select the protein family. The protein family name will then be listed in the "Proteins Available for Viewing" panel in the lower left area of the applet window.

In the Visualization panel 3808 on the right side of the applet window, the protein may be displayed in one of several formats. In the preferred embodiment, these formats include Description, Sequence, Structure, and Quality views, which are selected using the computer display mouse and the tabs along the top edge of the Visualization panel. The details of these display formats are the same as for the corresponding views described above in conjunction with Figures 11 through 30. As noted below, however, the alternative embodiment of Figure 36 uses a distributed network architecture for the database and for the user-to-database interface, and thereby permits simultaneous display of multiple proteins and associated data in the Visualization panel through EJB components.

The Visualization panel of Figure 38 also shows that various sub-panels 3810 may be displayed and manipulated. For example, in the lower right area of the Visualization panel, protein views may be selected to show Ramachandran plots, Balasubram plots, Hydrophobicity data, Strain plots, and Profiles plots. In addition, Sequence and Secondary Structure information may be shown simultaneously, as indicated by the sub-panel adjacent to the Profile Analysis view at the bottom right of the Visualization panel. The details of these display formats are the same as for the corresponding views described above in conjunction with Figures 11 through 30. As

noted below, however, the alternative embodiment of Figure 36 uses a distributed network architecture for the database and for the user-to-database interface and thereby permits simultaneous display of multiple proteins and associated data in the Visualization panel.

Figure 39 is a representation of the Application screen 3900 shown to a user at a client machine of the network system illustrated in Figure 36, illustrating the Search panel that is shown in the application window when a user selects that display tab from the Protein Selection display area on the left side of the display 3904. Figure 39 shows that a user enters a protein pattern identifier in a protein dialogue box 3910 of the display and then selects a Search display button 3912. The results of a search for the identified protein are then listed in a text window 3914, showing the protein families that matched the search string. Those skilled in the art will understand the protein pattern identifier notation used to identify protein families. Beneath the pattern search panel, the web client applet shows sequence information 3916 for a user-selected one of the search result proteins. As noted above, one of the proteins may then be selected for viewing in the Visualization area of the display screen, and multiple proteins may be selected for simultaneous display.

Figure 40 is a representation of the Application screen 4000 shown to a user at a client machine of the network system illustrated in Figure 36, showing the Data Mining panel selected using the display tab of the Protein Selection panel 4004 at the left side of the application window. Figure 40 shows that selection of the Data Mining tab generates a dialogue box 4010 in which a protein name is entered, and on which a similarity search

will be conducted. Figure 40 shows that a "Find All Similar Proteins" display button 4012 is provided. When the user selects the display button, the web client applet forwards the protein search identifier entered by the user in the dialogue box 4010 and provides it to the application server and then to the database management system. A database search is then conducted and a list of similar proteins is provided in the text window 4014 shown beneath the "Find" button.

From the list of Similar Proteins found from the database, a user can select one or more of the "Similar Proteins" for further information. For example, Figure 40 shows a panel 4016 below the "Find" box that is called "Similar Cores Between Selected Proteins", in which the target protein entered in the dialogue box is shown, followed by information for a selected one of the "Similar Proteins". In the "Proteins Available for Viewing" panel 4018, both the target protein (shown as "muscle type acylphosphatease") is listed and the "Similar Protein" (shown as "orphan nuclear receptor HMR") is shown. When these proteins are selected for viewing, as indicated by their entry in the "Proteins Available for Viewing" panel, their corresponding views are shown in the Visualization panel 4008.

Figure 41 is a representation of the Application screen 4100 shown to a user at a client machine of the network system illustrated in Figure 36, illustrating that simultaneous proteins can be shown in the Visualization panel 4108. In particular, Figure 41 shows Quality display features for two selected proteins. The two proteins are identified in the "Proteins Available for Viewing" panel 4108 at the lower left area of the Application screen, and their corresponding Ramachandran plots of the Quality view are

simultaneously shown in the visualization panel 4108, along with their respective Profile Analysis panels with Residue information. As noted above, it should be understood that all of the Description, Sequence, Structure, and Quality sub-panels and display views described above for the first embodiment are shown in the respective display views in the second embodiment. In the second embodiment of Figures 36 through 46, however, multiple proteins can be shown simultaneously, advantageously using the web client applets and the distributed architecture.

Figure 42 is a representation of the Application screen 4200 shown to a user at a client machine of the network system illustrated in Figure 36, illustrating the simultaneous display of protein Structure features for two selected proteins in the Visualization panel 4208 of the application window. Moreover, Figure 42 shows that the applet panels can be resized in accordance with known window programming techniques, thereby making it possible to devote greater screen area to panels of greater interest. Thus, in Figure 42, the visual display Structure panel of the Visualization area has been resized to occupy a greater portion of the user's display screen as compared with the visualization panel of Figures 38 through 41. Accordingly, the Protein Selection area occupies less display screen area (compare, for example, with Figure 39). It should be apparent to those skilled in the art that the resizing is achieved with the left and right buttons of a display mouse, appropriately dragging the display cursor for the desired resizing. Again, details of the information shown in the Structure display of the Visualization panel are the same as those described above for the first embodiment. The second embodiment, however, permits simultaneous display of multiple proteins.

Figure 43 is a representation 4300 of the Application screen of Figure 42, this time showing the drop down menu 4320 for selection of Display features. More particularly, Figure 43 of the Display item of the applet menu bar shows that a user may select between a Skeleton display format, a Ball & Sticks format, a Spacefill format, and a VDW Dot Skeleton display format for the Visualization panel. These display formats will be known to those skilled in the art, so that such persons will understand what information will be shown in such display views without further explanation. Again, details of the information shown in the Display formats of the Visualization panel are the same as those described above for corresponding Display views of the first embodiment.

Figure 44 is a representation 4400 of the Application screen of Figure 42, showing the drop down menu 4420 for selection of Options for atoms to be viewed. More particularly, Figure 44 of the Options item of the applet menu bar shows that a user may select between a "C-Alpha Atoms" view, a "Main Chain Atoms" view, and an "All Atoms" view for the Visualization panel. These view formats will be known to those skilled in the art, so that such persons will understand what information will be shown in such display views without further explanation. Again, details of the information shown in the Options formats of the Visualization panel are the same as those described above for corresponding views of the first embodiment.

Figure 45 is a representation of the Application screen 4500 shown to a user at a client machine of the network system illustrated in Figure 36, illustrating simultaneous Sequence display features for two selected proteins in the Visualization panel 4508. Figure 45 shows the Visualization panel in the resized condition first shown in Figure 42,

illustrating sequence information for multiple proteins. As before, the proteins are selected through the "Proteins Available for Viewing" panel 4508 at the lower left of the display.

Figure 46 is a graphical representation of the database objects for the database design of the system illustrated in Figure 36. As with the first embodiment, the system of the second embodiment shown in Figure 36 is implemented using object oriented programming techniques in which data objects are organized into classes, each of which is characterized by attributes that specify parameters of the class and methods or processes that specify behaviors of the class.

More particularly, Figure 46 shows that the database design of the second embodiment includes a Residue class, an Atom class, a Domain class, and an Active Sites class. The database also includes a Protein class, a Protein Sequence class, a Sequence Link class, a Chain class, Secondary Structure class, VRMLURL class, Protein Segment class, Protein Search class, Transformation class, an Alignment class, an Alignment Residue class, and an Elements class. The database design also includes a Subfamily class, a Subfamily Proteins class, and a Family class. Finally, the database includes a Useraccess class, a UserAccessProteins class, a Feedback class, and a BugGroup class. Attributes of the database classes are shown in the class boxes of Figure 46. These classes store attribute data values and specify class behaviors to provide the functionality described herein.

For example, the Residue class stores parameters to produce a protein residue display with features such as illustrated in connection with the first embodiment (Figures

1 through 35) and the second embodiment (Figures 36 through 46). That is, the Residue class contains the information typically needed to specify a residue display, which will be apparent to those skilled in the art without further explanation. Similarly, the Atom class contains information needed to specify display of a protein atom, the Domain class contains information needed to specify a protein domain, and the Active Sites class contains information needed to specify the active sites of a protein for display. Those skilled in the art will understand the information needed to specify display of such features, which are like those specified by the first embodiment described above for corresponding displays in conjunction with Figures 1 through 35. Those skilled in the art will likewise appreciate the information needed for the other classes shown in Figure 46, in conjunction with the description of classes for the first embodiment and for display of the features described above.

Thus, the second embodiment illustrated in Figures 36 through 46 implements a bioinformatics database access system having a graphical user interface (GUI) that communicates over a shared network (such as the Internet) using graphical browsers to establish a communications session. Once communications with the database server are established, the GUI is provided through a platform-independent applet environment, such as provided with a Java programming environment. Thus, no hypertext mark-up language (HTML) page links are needed, and no common gateway interface (CGI) scripts are needed to exchange information and retrieve database information and create displays. Such processing, for example, also permits the simultaneous display of database information for more than one protein. Such programming also permits a

Protein Selection panel to be shown adjacent a Visualization panel, thereby permitting search and selection of proteins, followed by display of visualization data for such proteins, from the same display window. Thus, families of proteins can be shown along the left side of the display, while visualization displays of the proteins can be shown simultaneously along the right side of the display. This functionality is possible because, with the Java implementation, the same applet that communicates with the bioinformatics database also performs the visualization display tasks.

The present invention has been described above in terms of presently preferred embodiments so that an understanding of the present invention can be conveyed. There are, however, many configurations for protein database viewing systems not specifically described herein but with which the present invention is applicable. The present invention should therefore not be seen as limited to the particular embodiments described herein, but rather, it should be understood that the present invention has wide applicability with respect to molecular structure database viewing systems generally. All modifications, variations, or equivalent arrangements and implementations that are within the scope of the attached claims should therefore be considered within the scope of the invention.

CLAIMS

WE CLAIM:

1. A method of accessing molecular structure information over a computer network and graphically viewing the information, the method comprising:

receiving authorization information from a user and checking for authorization by that user to a database containing molecular structure information;

receiving information from a user identifying requested molecular structure information;

providing the user access authorization;

downloading multiple files containing the requested molecular structure information from a database server to the authorized user; and

viewing the downloaded files at the user with a graphical browser application program.

2. A method as defined in claim 1, wherein the database comprises a relational database containing molecular structure information for multiple protein structures.

3. A method as defined in claim 1, wherein providing user access authorization comprises checking user account information that permits user access to less than the entire database.

4. A method as defined in claim 3, wherein providing user access authorization comprises providing the user with a display of database portions to which the user has been granted authorization.

5. A method as defined in claim 4, further comprising displaying the entire available database and indicating those database portions to which the user has been granted authorization.

6. A method as defined in claim 1, wherein viewing includes executing helper applications to view different file formats from within the browser application program.

7. A method as defined in claim 6, further comprising displaying a molecular structure and displaying a sequence alignment view in an external frame that can be separately manipulated by the user and that interacts with the molecular structure display.

8. A method as defined in claim 7, further comprising displaying secondary structure elements with predetermined structural symbols.

9. A method as defined in claim 6, wherein the file formats include virtual reality formats.

10. A method as defined in claim 1, wherein providing the user access authorization occurs at a security server, and the step of downloading multiple files occurs from a database file server.

11. A method as defined in claim 1, wherein receiving information from a user and receiving multiple files comprises display of corresponding information in a single application window of a computer display.

12. A method as defined in claim 1, wherein receiving multiple files comprises display of visualization data for multiple proteins selected by the user.

13. A computer database system comprising:

a security server that receives user requests for user access authorization to a database containing molecular structure information and receives identification information from the user identifying requested molecular structure information, and then determines if user access authorization should be provided;

a database server that responds to a user access authorization by downloading multiple files containing the requested molecular structure information to the authorized user; and

a graphical browser application program that receives the downloaded files at the user and displays the information contained therein.

14. A system as defined in claim 13, wherein the database comprises a relational database containing molecular structure information for multiple protein structures.

15. A system as defined in claim 13, wherein the security server provides user access authorization by checking user account information that permits user access to less than the entire database.

16. A system as defined in claim 15, wherein the security server provides user access authorization by providing the user with a display of database portions to which the user has been granted authorization.

17. A system as defined in claim 16, wherein the security server displays the entire available database to the user prior to downloading and indicates those database portions to which the user has been granted authorization.

18. A system as defined in claim 13, wherein the user graphical browser application program executes helper applications to view different file formats from within the browser application program.

19. A system as defined in claim 18, wherein the downloaded files received by the user graphical browser application program permit it to display a molecular structure and to display a sequence alignment view in an external frame that can be separately manipulated by the user and that interacts with the molecular structure display.

20. A system as defined in claim 19, wherein the graphical browser application program displays secondary structure elements with predetermined structural symbols.

21. A system as defined in claim 18, wherein the file formats include virtual reality formats.

22. A system as defined in claim 13, wherein the security server and database server are separate computers connected over a network.

23. A method of operating a server for controlling access to molecular structure information over a computer network, the method comprising:

receiving authorization information from a user and checking for authorization by that user to a database containing molecular structure information;

receiving information from a user identifying requested molecular structure information; and

granting the user authorization for access if it is determined that the user should be provided with access authorization to permit downloading multiple files containing the requested molecular structure information from a database server to the authorized user, where the user can view the downloaded files with a graphical browser application program.

24. A method as defined in claim 23, wherein the database comprises a relational database containing molecular structure information for multiple protein structures.

25. A method as defined in claim 23, wherein the step of granting user access authorization comprises checking user account information that permits user access to less than the entire database.

26. A method as defined in claim 25, wherein the step of granting user access authorization comprises providing the user with a display of database portions to which the user has been granted authorization.

27. A method as defined in claim 26, further comprising the step of permitting display of the entire available database and indicating those database portions to which the user has been granted authorization.

28. A method as defined in claim 23, wherein the granted authorization permits downloading of different file formats that are viewed from within the browser application program with helper applications.

29. A method as defined in claim 28, wherein the security server permits downloading of files containing data that permit displaying a sequence alignment view in an external frame of the browser application program such that it can be separately manipulated by the user and interacts with the molecular structure display.

30. A method as defined in claim 29, further comprising the step of displaying secondary structure elements with predetermined structural symbols.

31. A method as defined in claim 28, wherein the file formats include virtual reality formats.

32. A method as defined in claim 23, wherein the step of providing the user access authorization occurs at a security server, and the step of downloading multiple files occurs from a database file server.

33. A method of operating a database server for providing molecular structure information over a computer network, the method comprising:

receiving a user authorization following a security authorization check that grants a user access to a database containing molecular structure information;

receiving information identifying requested molecular structure information; and

downloading multiple files containing the requested molecular structure information to the authorized user, where the user can view the downloaded files with a graphical browser application program.

34. A method as defined in claim 33, wherein the database comprises a relational database containing molecular structure information for multiple protein structures.

35. A method as defined in claim 33, wherein the granted authorization permits downloading of different file formats that are viewed from within the browser application program with helper applications.

36. A method as defined in claim 35, wherein the downloaded files include data that permits displaying a sequence alignment view in an external frame of the browser application program such that it can be separately manipulated by the user and interacts with the molecular structure display.

37. A method as defined in claim 36, further comprising the step of displaying secondary structure elements with predetermined structural symbols.

38. A method as defined in claim 35, wherein the file formats include virtual reality formats.

39. A method of operating a user computer for accessing molecular structure information over a computer network, the method comprising:

providing authorization information to a security server to request authorization to a database containing molecular structure information and receiving an authorization response;

providing information that identifies requested molecular structure information;

receiving multiple files in an authorized download operation, the downloaded files containing the requested molecular structure information; and

viewing the downloaded files with a graphical browser application program.

40. A method as defined in claim 39, wherein the database comprises a relational database containing molecular structure information for multiple protein structures.

41. A method as defined in claim 39, wherein the authorization information includes information sufficient for the security server to check user account information that permits user access to less than the entire database.

42. A method as defined in claim 41, wherein the authorization response includes information that provides a display of database portions to which authorization has been granted.

43. A method as defined in claim 42, wherein the authorization response further includes information that provides a display of the entire available database and indicates those database portions to which authorization has been granted.

44. A method as defined in claim 39, wherein the granted authorization permits downloading of different file formats that are viewed from within the browser application program with helper applications.

45. A method as defined in claim 44, wherein the downloaded files contain data that permits displaying a sequence alignment view in an external frame of the browser application program such that it can be separately manipulated by the user and interacts with the molecular structure display.

46. A method as defined in claim 45, further comprising the step of displaying secondary structure elements with predetermined structural symbols.

47. A method as defined in claim 39, wherein the file formats include virtual reality formats.

48. A method as defined in claim 39, wherein the downloaded multiple files are received from a database file server.

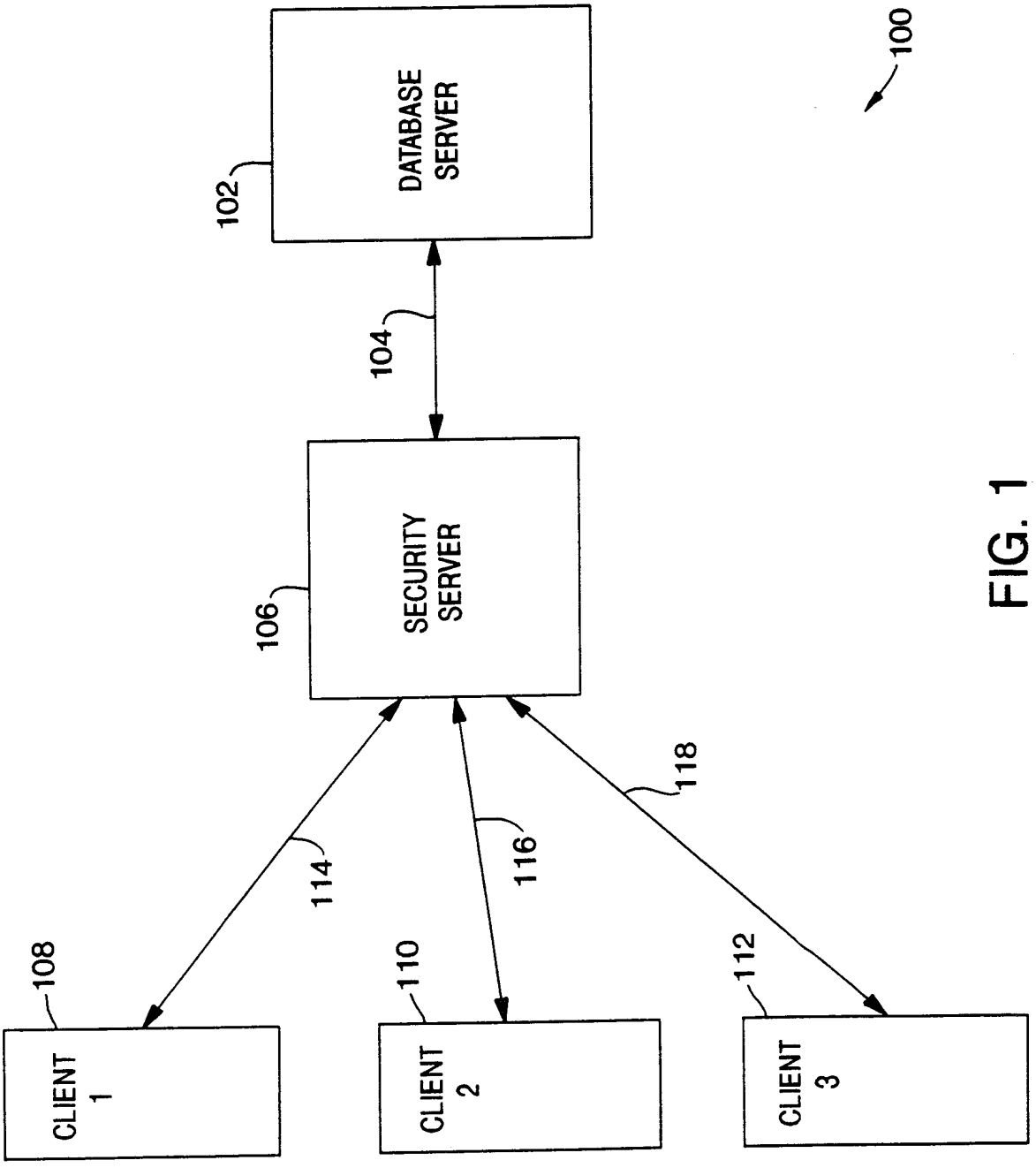

100
DATABASE SERVER
102
104
SECURITY SERVER
106
114
116
118
108
110
112
CLIENT 1
CLIENT 2
CLIENT 3

FIG. 1

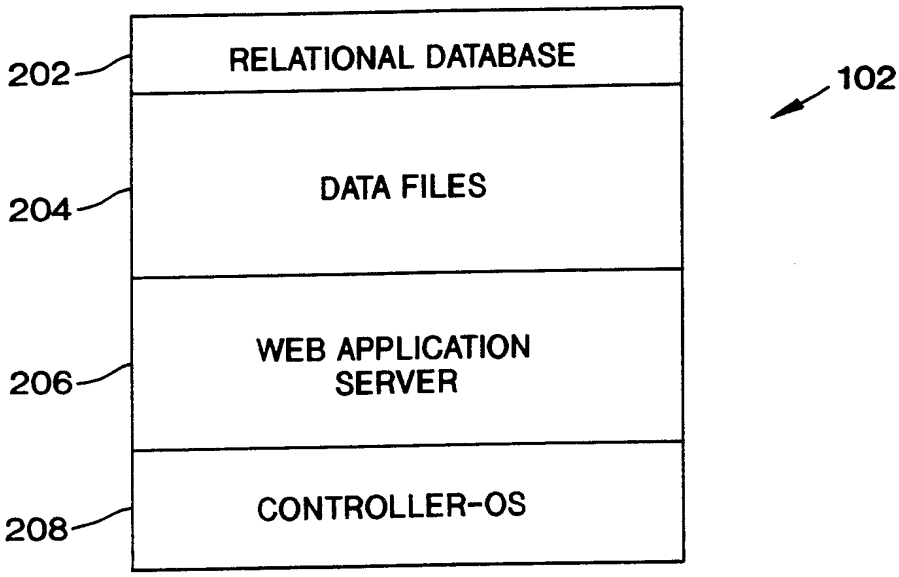

RELATIONAL DATABASE
DATA FILES
WEB APPLICATION SERVER
CONTROLLER-OS
202
204
206
208
102

FIG. 2

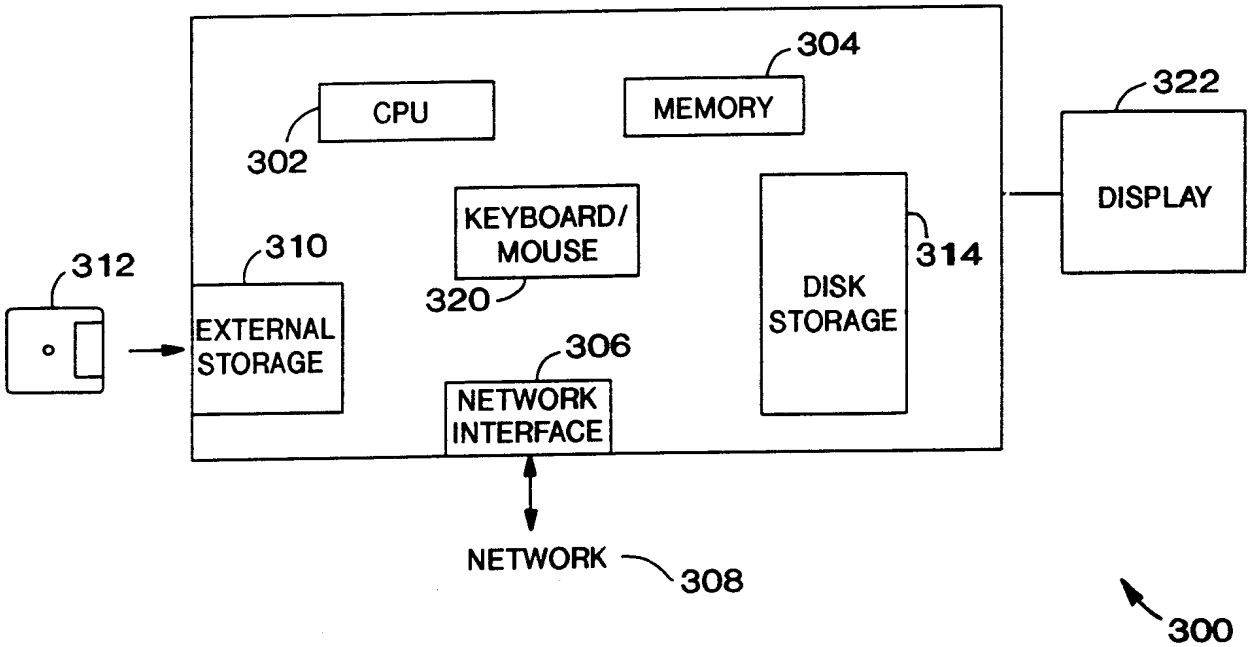

CPU
302
MEMORY
304
KEYBOARD/ MOUSE
320
DISK STORAGE
314
DISPLAY
322
EXTERNAL STORAGE
310
312
NETWORK INTERFACE
306
NETWORK
308
300

FIG. 3

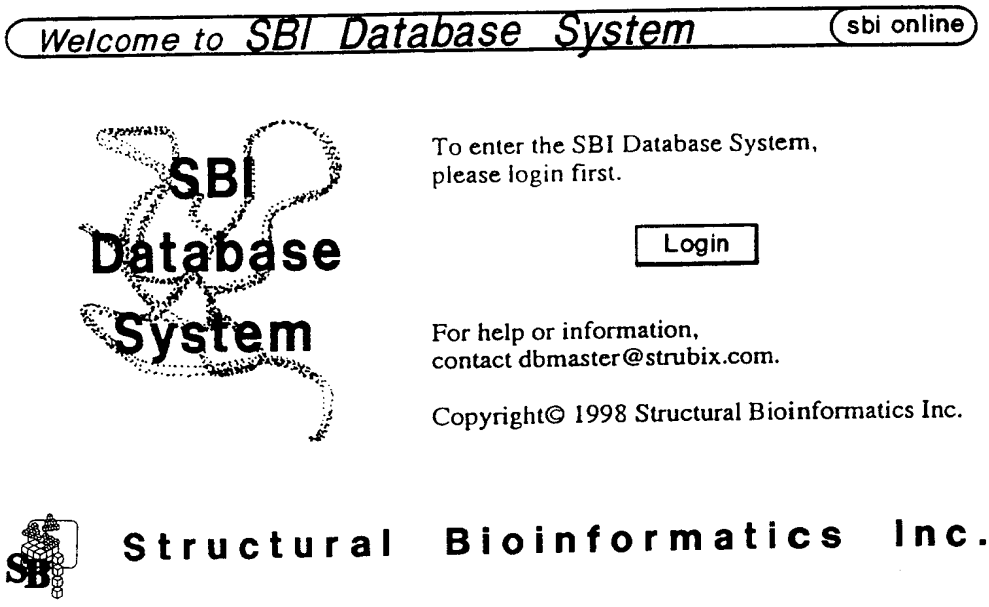

Welcome to SBI Database System
sbi online
SBI
Database
System
To enter the SBI Database System,
please login first.
Login
For help or information,
contact dbmaster@strubix.com.
Copyright© 1998 Structural Bioinformatics Inc.
Structural Bioinformatics Inc.

FIG. 4

SBI Database System sbi online

Information Box

Please click on one of the following utilities.

Click on one.

SBdBase

Comprehensive database of protein structural models derived from genomic sequences

SBdBase Plus

Extension of SBdBase including drug design

SVdBase

Comprehensive database of protein structural models derived for drug-target specific genetic polymorphisms

CombiLib

Virtual library of multi-conformer combinatorial chemicals

SBdBase

SBdBase Plus

SVdBase

CombiLib

For further information or assistance, email dbmaster@strubix.com.

FIG. 5

SBI Database System sbi online

Information Box
Please click on one of the following utilities.
Click on one.
SBdBase Index
Select protein from SBdBase Index
SBdBase Query
Submit query to SBdBase

SBdBase Index

Query Submission

For further information or assistance, email dbmaster@strubix.com.

FIG. 6

- Fructose-1
- Galaptin Family
- Glucose-6-phosphate Dehydrogenase Family
- Glutamate Dehydrogenase Family
- Glyceraldehyde-3-phosphate Dehydrogenase Family
- Glycoprotein Hormone Family
- Glycosyl Hydrolase Family
- Helical Cytokine Family
- Hemopexin-Domain Family
- Immunoglobulin Family
- Inosinate Dehydrogenase Family
- Inositol Phosphatase Family
- Integrin-alpha Family
- Interferon Family
- Interleukin-1 Family
- Isocitrate/Isopropylmalate Dehydrogenase Family
- Lactate Dehydrogenase Family
- Leukemia Inhibitory Factor/Oncostatin M Family
- MHC Class I Family
- MHC Class II Family
- MYC Helix-loop-helix Family
- Malate Dehydrogenase Family
- Matrix Metalloprotease Family
- Metallothionein Family
- Migratory Inhibitory Factor Family
- Natriuretic Peptide Family
- Nerve Growth Factor Family
- PTI Kunitz Family
- Pectate Lyase Family
- Phosphoglycerate Mutase Family
- Phospholipase Family

FIG. 7A

- Platelet-derived Growth Factor Family
- Protein Kinase Family
- SCP Extracellular Family
- SH3 Family
- ST Phosphatase Family
- Serine Protease Family
- Somatotropin Family
- TGF-beta Family
- Thy Phosphorylase Family
- Trefoil Family
- Tumor Necrosis Factor Family
- Tumor Necrosis Factor Receptor Family
- Ubiquitin Family
- bZIP Transcription Factor Family

FIG. 7B

SBI Database System (sbi online)

Information Box
To do a query, fill out the form below and click on the SUBMIT QUERY button.
Fill out the query form.
ID Number
SBdBase identification number for a protein
e.g. SBI12345
Protein Name
Full or partial name for a protein
e.g. Bak or regulator
Species
Species of interest
e.g. Human
Gene
Name of the gene that encodes for a protein
e.g. MMP1
Accession Number
SwissProt accession number or code for a protein
e.g. P30885 or AP1_HUMAN
Disease
Disease state for which protein(s) are to be retrieved
e.g. agammaglobulinemia
Function
Function for which protein(s) are to be retrieved
e.g. differentiation
Family
Family of proteins to be retrieved
e.g. peptidase
Full Text
Search the full text of each entry

SBdBase ID Number
Protein Name
Species
Gene
Accession Number
Disease
Function
Family
Full Text Search

AND OR

Submit Query Clear Form

Please click on the SUBMIT QUERY button only once.

For further information or assistance, email dbmaster@strubix.com.

FIG. 8

SBI Database System (sbi online)

Protein Name	SBdBase ID	PDB Entry	SBdBase
HEPATOCYTE GROWTH FACTOR RECEPTOR PRECURSOR (MET PROTO-ONCOGENE TYROSINE KINASE) (EC 2.7.1.112) (HGF-SF RECEPTOR).	SBI23580		SBbBase
ARYLAMINE N-ACETYLTRANSFERASE, POLYMORPHIC (EC 2.3.1.5) (PNAT).	SBI9549		
BREAST CANCER TYPE 1 SUSCEPTIBILITY PROTEIN.	SBI9666		
ALPHA-1 CATENIN (CADHERIN-ASSOCIATED PROTEIN) (ALPHA E-CATENIN).	SBI10091		
ALPHA-CATENIN (CADHERIN-ASSOCIATED PROTEIN) (ALPHA E-CATENIN).	SBI10093		
DIHYDROPYRIMIDINE DEHYDROGENASE (NADP+) PRECURSOR (EC 1.3.1.2) (DPD) (DIHYDROURACIL DEHYDROGENASE) (DIHYDROTHYMINE DEHYDROGENASE).	SBI10223		
FRAGILE HISTIDINE TRIAD PROTEIN.	SBI10445	PDB	
PARANEOPLASTIC ENCEPHALOMYELITIS ANTIGEN HUD (HU-ANTIGEN D).	SBI10865		
MUTL PROTEIN HOMOLOG 1 (DNA MISMATCH REPAIR PROTEIN MLH1).	SBI11500		
DNA MISMATCH REPAIR PROTEIN MSH2.	SBI11536		
ONCONEURAL VENTRAL ANTIGEN-1 (NOVA-1) (PARANEOPLASTIC RI ANTIGEN).	SBI11685		
PROHIBITIN.	SBI11920		
PMS1 PROTEIN HOMOLOG 1 (DNA MISMATCH REPAIR PROTEIN PMS1).	SBI11969		
PMS1 PROTEIN HOMOLOG 2 (DNA MISMATCH REPAIR PROTEIN PMS2).	SBI11970		
PROTEIN-TYROSINE PHOSPHATASE G1 (EC 3.1.3.48) (PTPG1).	SBI12084		
RETINOBLASTOMA-ASSOCIATED PROTEIN (PP110) (P105-RB) (RB).	SBI12157		
TUMOR NECROSIS FACTOR PRECURSOR (TNF-ALPHA) (CACHECTIN).	SBI12647	PDB	
TUBERIN (TUBEROUS SCLEROSIS 2 HOMOLOG PROTEIN).	SBI24366		
TUMOR NECROSIS FACTOR PRECURSOR (TNF-ALPHA) (CACHECTIN).	SBI24326		

FIG. 9A

RECOVERIN (CANCER ASSOCIATED RETINOPATHY PROTEIN) (CAR PROTEIN).	SBI23914
THYROTROPIN RECEPTOR PRECURSOR (TSH-R) (THYROID STIMULATING HORMONE RECEPTOR).	SBI12716
VON HIPPEL-LINDAU DISEASE TUMOR SUPPRESSOR (G7 PROTEIN) (FRAGMENT).	SBI12863
DNA-REPAIR PROTEIN COMPLEMENTING XP-A CELLS (XERODERMA PIGMENTOSUM GROUP A COMPLEMENTING PROTEIN).	SBI12900
DNA-REPAIR PROTEIN COMPLEMENTING XP-C CELLS (XERODERMA PIGMENTOSUM GROUP C COMPLEMENTING PROTEIN) (P125).	SBI12902
DNA-REPAIR PROTEIN COMPLEMENTING XP-D CELLS (XERODERMA PIGMENTOSUM GROUP D COMPLEMENTING PROTEIN) (DNA EXCISION REPAIR PROTEIN ERCC-2).	SBI12904
DNA-REPAIR PROTEIN COMPLEMENTING XP-F CELL (XERODERMA PIGMENTOSUM GROUP F COMPLEMENTING PROTEIN) (DNA EXCISION REPAIR PROTEIN ERCC-4).	SBI12905
CELLULAR TUMOR ANTIGEN P53.	SBI18260
TUMOR NECROSIS FACTOR PRECURSOR (TNF-ALPHA) (CACHECTIN).	SBI18684
CELLULAR TUMOR ANTIGEN P53.	SBI19105
CELLULAR TUMOR ANTIGEN P53.	SBI21125
TUMOR NECROSIS FACTOR PRECURSOR (TNF-ALPHA) (CACHECTIN).	SBI21634

FIG. 9B

SBI Database System (sbi online)

Information Box

To do a query, fill out the form below and click on the SUBMIT QUERY button.

Fill out the query form.

ID Number

SBdBase identification number for a protein

e.g. SBI12345

Protein Name

Full or partial name for a protein

e.g. Bak or regulator

Species

Species of interest

e.g. Human

Gene

Name of the gene that encodes for a protein

e.g. MMP1

Accession Number

SwissProt accession number or code for a protein

e.g. P30885 or AP1_HUMAN

Disease

Disease state for which protein(s) are to be retrieved

e.g. agammaglobulinemia

Function

Function for which protein(s) are to be retrieved

e.g. differentiation

Family

Family of proteins to be retrieved

e.g. peptidase

Full Text

Search the full text of each entry

SBdBase ID Number
Protein Name
Species
Gene
Accession Number
Disease cancer
Function
Family
Full Text Search AND OR

Submit Query Clear Form

Please click on the SUBMIT QUERY button only once.

FIG. 10

SBI Database System sbi online

SBdBase ID	SBI9326
Protein Name	6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING (EC 1.1.1.44).
Species	HUMAN
Gene	PGD
Keywords	OXIDOREDUCTASE , PENTOSE SHUNT , NADP ,
Disease Information	None
Function Information	None
Family Class	TO OTHER PROKARYOTIC AND EUKARYOTIC 6-PHOSPHOGLUCONATE DEHYDROGENASES.
SWISSPROT Accession Number	P52209 (6PGD_HUMAN)
EC Number	EC 1.1.1.44

For more information on this protein, click on any of the selections below.

SBdBase SwissProt GenBank PIR

New Query

For further information or assistance, email dbmaster@strubix.com

FIG. 11

```
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: P 1
REMARK 290
REMARK 290      SYMOP   SYMMETRY
REMARK 290     NNNMMM   OPERATOR
REMARK 290       1555   X,Y,Z
REMARK 290
REMARK 290     WHERE NNN -> OPERATOR NUMBER
REMARK 290           MMM -> TRANSLATION VECTOR
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS
REMARK 290 THE FOLLOWING TRANSFORMATIONS OPERATE ON THE ATOM/HETATM
REMARK 290 RECORDS IN THIS ENTRY TO PRODUCE CRYSTALLOGRAPHICALLY
REMARK 290 RELATED MOLECULES.
REMARK 290   SMTRY1   1  1.000000  0.000000  0.000000        0.00000
REMARK 290   SMTRY2   1  0.000000  1.000000  0.000000        0.00000
REMARK 290   SMTRY3   1  0.000000  0.000000  1.000000        0.00000
REMARK 290
REMARK 290 REMARK: NULL
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: COVALENT BOND LENGTHS
REMARK 500
REMARK 500 THE STEREOCHEMICAL PARAMETERS OF THE FOLLOWING RESIDUES
REMARK 500 HAVE VALUES WHICH DEVIATE FROM EXPECTED VALUES BY MORE
REMARK 500 THAN 4*RMSD AND BY MORE THAN 0.150 ANGSTROMS (M=MODEL
REMARK 500 NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE
REMARK 500 NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT: (10X,I3,1X,A3,1X,A1,I4,A1,1X,2(A4,A1,3X),12X,F5.3)
REMARK 500
REMARK 500 EXPECTED VALUES: ENGH AND HUBER, 1991
REMARK 500
REMARK 500  M RES CSSEQI ATM1    RES CSSEQI ATM2    DEVIATION
REMARK 500  0 MET B  87   SD     MET B  87   CE       0.164
REMARK 650
REMARK 650 HELIX
REMARK 650 DETERMINATION METHOD: AUTHOR-DETERMINED
REMARK 999
REMARK 999 SEQUENCE
REMARK 999 1AO6  A    SWS      P02768        1 -    28 NOT IN ATOMS LIST
REMARK 999 1AO6  A    SWS      P02768      607 -   609 NOT IN ATOMS LIST
REMARK 999 1AO6  B    SWS      P02768        1 -    28 NOT IN ATOMS LIST
REMARK 999 1AO6  B    SWS      P02768      607 -   609 NOT IN ATOMS LIST
DBREF  1AO6 A    5   582  SWS    P02768   ALBU_HUMAN      29    606
DBREF  1AO6 B    5   582  SWS    P02768   ALBU_HUMAN      29    606
SEQRES   1 A  585  ASP ALA HIS LYS SER GLU VAL ALA HIS ARG PHE LYS ASP
SEQRES   2 A  585  LEU GLY GLU GLU ASN PHE LYS ALA LEU VAL LEU ILE ALA
SEQRES   3 A  585  PHE ALA GLN TYR LEU GLN GLN CYS PRO PHE GLU ASP HIS
SEQRES   4 A  585  VAL LYS LEU VAL ASN GLU VAL THR GLU PHE ALA LYS THR
SEQRES   5 A  585  CYS VAL ALA ASP GLU SER ALA GLU ASN CYS ASP LYS SER
SEQRES   6 A  585  LEU HIS THR LEU PHE GLY ASP LYS LEU CYS THR VAL ALA
SEQRES   7 A  585  THR LEU ARG GLU THR TYR GLY GLU MET ALA ASP CYS CYS
SEQRES   8 A  585  ALA LYS GLN GLU PRO GLU ARG ASN GLU CYS PHE LEU GLN
SEQRES   9 A  585  HIS LYS ASP ASP ASN PRO ASN LEU PRO ARG LEU VAL ARG
SEQRES  10 A  585  PRO GLU VAL ASP VAL MET CYS THR ALA PHE HIS ASP ASN
SEQRES  11 A  585  GLU GLU THR PHE LEU LYS LYS TYR LEU TYR GLU ILE ALA
SEQRES  12 A  585  ARG ARG HIS PRO TYR PHE TYR ALA PRO GLU LEU LEU PHE
SEQRES  13 A  585  PHE ALA LYS ARG TYR LYS ALA ALA PHE THR GLU CYS CYS
SEQRES  14 A  585  GLN ALA ALA ASP LYS ALA ALA CYS LEU LEU PRO LYS LEU
SEQRES  15 A  585  ASP GLU LEU ARG ASP GLU GLY LYS ALA SER SER ALA LYS
SEQRES  16 A  585  GLN ARG LEU LYS CYS ALA SER LEU GLN LYS PHE GLY GLU
SEQRES  17 A  585  ARG ALA PHE LYS ALA TRP ALA VAL ALA ARG LEU SER GLN
SEQRES  18 A  585  ARG PHE PRO LYS ALA GLU PHE ALA GLU VAL SER LYS LEU
SEQRES  19 A  585  VAL THR ASP LEU THR LYS VAL HIS THR GLU CYS CYS HIS
SEQRES  20 A  585  GLY ASP LEU LEU GLU CYS ALA ASP ASP ARG ALA ASP LEU
```

FIG. 12A

```
SEQRES   43 B  585  VAL MET ASP ASP PHE ALA ALA PHE VAL GLU LYS CYS CYS
SEQRES   44 B  585  LYS ALA ASP ASP LYS GLU THR CYS PHE ALA GLU GLU GLY
SEQRES   45 B  585  LYS LYS LEU VAL ALA ALA SER GLN ALA ALA LEU GLY LEU
FORMUL    3  HOH    *7(H2 O1)
HELIX     1 A1A SER A    5  ASP A   13  1                                   9
HELIX     2 A2A GLU A   16  LEU A   31  1                                  16
HELIX     3 A3A PRO A   35  ASP A   56  1                                  22
HELIX     4 A4A SER A   65  THR A   76  1                                  12
HELIX     5 A5A ALA A   88  ALA A   92  1                                   5
HELIX     6 A6A GLU A   95  HIS A  105  1                                  11
HELIX     7 B1A GLU A  119  HIS A  128  1                                  10
HELIX     8 B2A GLU A  131  HIS A  146  1                                  16
HELIX     9 B3A TYR A  150  GLN A  170  1                                  21
HELIX    10 B4A ASP A  173  LYS A  195  1 B4A IS FUSED WITH C1A.            23
HELIX    11 C1A GLN A  196  LYS A  205  1                                  10
HELIX    12 C2A GLU A  208  PHE A  223  1                                  16
HELIX    13 C3A GLU A  227  HIS A  247  1                                  21
HELIX    14 C4A LEU A  250  GLN A  268  1                                  19
HELIX    15 C5A LEU A  275  GLU A  280  1                                   6
HELIX    16 C6A LEU A  283  GLU A  292  1                                  10
HELIX    17 D1A ASP A  314  GLU A  321  1                                   8
HELIX    18 D2A ASP A  324  HIS A  338  1                                  15
HELIX    19 D3A SER A  342  ALA A  362  1                                  21
HELIX    20 D4A ASP A  365  GLU A  383  1 D4A IS FUSED WITH E1A.            19
HELIX    21 E1A PRO A  384  LEU A  398  1                                  15
HELIX    22 E2A TYR A  401  VAL A  415  1                                  15
HELIX    23 E3A SER A  419  CYS A  438  1                                  20
HELIX    24 E4A LYS A  444  THR A  467  1                                  24
HELIX    25 E5A SER A  470  GLU A  479  1                                  10
HELIX    26 E6A ASN A  483  LEU A  491  1                                   9
HELIX    27 F1A HIS A  510  THR A  515  1                                   6
HELIX    28 F2A GLU A  518  LYS A  536  1                                  19
HELIX    29 F3A GLU A  542  LYS A  560  1                                  19
HELIX    30 F4A GLU A  565  ALA A  582  1                                  18
HELIX    31 A1B SER B   5  ASP B   13  1                                   9
HELIX    32 A2B GLU B  16  LEU B   31  1                                  16
HELIX    33 A3B PRO B  35  ASP B   56  1                                  22
HELIX    34 A4B SER B  65  THR B   76  1                                  12
HELIX    35 A5B ALA B  88  ALA B   92  1                                   5
HELIX    36 A6B GLU B  95  HIS B  105  1                                  11
HELIX    37 B1B GLU B 119  HIS B  128  1                                  10
HELIX    38 B2B GLU B 131  HIS B  146  1                                  16
HELIX    39 B3B TYR B 150  GLN B  170  1                                  21
HELIX    40 B4B ASP B 173  LYS B  195  1 B4B IS FUSED WITH C1B.            23
HELIX    41 C1B GLN B 196  LYS B  205  1                                  10
HELIX    42 C2B GLU B 208  PHE B  223  1                                  16
HELIX    43 C3B GLU B 227  HIS B  247  1                                  21
HELIX    44 C4B LEU B 250  GLN B  268  1                                  19
HELIX    45 C5B LEU B 275  GLU B  280  1                                   6
HELIX    46 C6B LEU B 283  GLU B  292  1                                  10
HELIX    47 D1B ASP B 314  GLU B  321  1                                   8
HELIX    48 D2B ASP B 324  HIS B  338  1                                  15
HELIX    49 D3B SER B 342  ALA B  362  1                                  21
HELIX    50 D4B ASP B 365  GLU B  383  1 D4B IS FUSED WITH E1B.            19
HELIX    51 E1B PRO B 384  LEU B  398  1                                  15
HELIX    52 E2B TYR B 401  VAL B  415  1                                  15
HELIX    53 E3B SER B 419  CYS B  438  1                                  20
HELIX    54 E4B LYS B 444  THR B  467  1                                  24
HELIX    55 E5B SER B 470  GLU B  479  1                                  10
HELIX    56 E6B ASN B 483  LEU B  491  1                                   9
HELIX    57 F1B HIS B 510  THR B  515  1                                   6
HELIX    58 F2B GLU B 518  LYS B  536  1                                  19
HELIX    59 F3B GLU B 542  LYS B  560  1                                  19
HELIX    60 F4B GLU B 565  ALA B  582  1                                  18
SSBOND    1 CYS A   53    CYS A   62
SSBOND    2 CYS A   75    CYS A   91
SSBOND    3 CYS A   90    CYS A  101
```

FIG. 12B

```
SSBOND   4 CYS A  124    CYS A  169
SSBOND   5 CYS A  168    CYS A  177
SSBOND   6 CYS A  200    CYS A  246
SSBOND   7 CYS A  245    CYS A  253
SSBOND   8 CYS A  265    CYS A  279
SSBOND   9 CYS A  278    CYS A  289
SSBOND  10 CYS A  316    CYS A  361
SSBOND  11 CYS A  360    CYS A  369
SSBOND  12 CYS A  392    CYS A  438
SSBOND  13 CYS A  437    CYS A  448
SSBOND  14 CYS A  461    CYS A  477
SSBOND  15 CYS A  476    CYS A  487
SSBOND  16 CYS A  514    CYS A  559
SSBOND  17 CYS A  558    CYS A  567
SSBOND  18 CYS B   53    CYS B   62
SSBOND  19 CYS B   75    CYS B   91
SSBOND  20 CYS B   90    CYS B  101
SSBOND  21 CYS B  124    CYS B  169
SSBOND  22 CYS B  168    CYS B  177
SSBOND  23 CYS B  200    CYS B  246
SSBOND  24 CYS B  245    CYS B  253
SSBOND  25 CYS B  265    CYS B  279
SSBOND  26 CYS B  278    CYS B  289
SSBOND  27 CYS B  316    CYS B  361
SSBOND  28 CYS B  360    CYS B  369
SSBOND  29 CYS B  392    CYS B  438
SSBOND  30 CYS B  437    CYS B  448
SSBOND  31 CYS B  461    CYS B  477
SSBOND  32 CYS B  476    CYS B  487
SSBOND  33 CYS B  514    CYS B  559
SSBOND  34 CYS B  558    CYS B  567
CISPEP   1 GLU A   95    PRO A   96          0        2.73
CISPEP   2 GLU B   95    PRO B   96          0        2.55
CRYST1   59.680   96.980   59.720  91.07 103.50  75.08 P 1           2
ORIGX1      1.000000  0.000000  0.000000        0.00000
ORIGX2      0.000000  1.000000  0.000000        0.00000
ORIGX3      0.000000  0.000000  1.000000        0.00000
SCALE1      0.016756 -0.004465  0.004224        0.00000
SCALE2      0.000000  0.010671 -0.000471        0.00000
SCALE3      0.000000  0.000000  0.017237        0.00000
ATOM      1  N   SER A   5      56.653  51.017  34.141  1.00 33.61           N
ATOM      2  CA  SER A   5      56.672  50.186  32.893  1.00 30.65           C
ATOM      3  C   SER A   5      55.611  49.138  33.086  1.00 30.62           C
ATOM      4  O   SER A   5      55.776  48.267  33.935  1.00 31.05           O
ATOM      5  CB  SER A   5      58.023  49.508  32.732  1.00 29.65           C
ATOM      6  OG  SER A   5      58.239  49.111  31.393  1.00 28.64           O
ATOM      7  N   GLU A   6      54.513  49.256  32.336  1.00 30.95           N
ATOM      8  CA  GLU A   6      53.399  48.316  32.414  1.00 30.41           C
ATOM      9  C   GLU A   6      53.710  46.993  31.698  1.00 29.24           C
ATOM     10  O   GLU A   6      53.409  45.928  32.221  1.00 31.26           O
ATOM     11  CB  GLU A   6      52.126  48.958  31.882  1.00 31.33           C
ATOM     12  CG  GLU A   6      50.890  48.067  32.001  1.00 35.35           C
ATOM     13  CD  GLU A   6      50.377  47.881  33.425  1.00 37.15           C
ATOM     14  OE1 GLU A   6      51.005  48.412  34.373  1.00 36.43           O
ATOM     15  OE2 GLU A   6      49.328  47.199  33.581  1.00 37.61           O
ATOM     16  N   VAL A   7      54.398  47.060  30.567  1.00 27.64           N
ATOM     17  CA  VAL A   7      54.785  45.871  29.816  1.00 27.81           C
ATOM     18  C   VAL A   7      55.810  45.022  30.581  1.00 29.85           C
ATOM     19  O   VAL A   7      55.941  43.817  30.341  1.00 31.73           O
ATOM     20  CB  VAL A   7      55.406  46.243  28.444  1.00 27.31           C
ATOM     21  CG1 VAL A   7      56.899  46.602  28.576  1.00 26.53           C
ATOM     22  CG2 VAL A   7      55.267  45.105  27.481  1.00 29.58           C
ATOM     23  N   ALA A   8      56.576  45.684  31.451  1.00 29.97           N
ATOM     24  CA  ALA A   8      57.619  45.075  32.258  1.00 28.93           C
ATOM     25  C   ALA A   8      56.965  44.373  33.429  1.00 30.57           C
ATOM     26  O   ALA A   8      57.379  43.274  33.806  1.00 31.75           O
ATOM     27  CB  ALA A   8      58.566  46.146  32.774  1.00 27.91           C
```

FIG. 12C

ATOM	3713	CG2	VAL	A	469	19.971	40.969	6.434	1.00	41.70	C
ATOM	3714	N	SER	A	470	16.483	41.447	9.465	1.00	42.33	N
ATOM	3715	CA	SER	A	470	15.124	41.227	9.974	1.00	43.02	C
ATOM	3716	C	SER	A	470	14.890	42.022	11.238	1.00	42.51	C
ATOM	3717	O	SER	A	470	15.620	41.896	12.209	1.00	40.52	O
ATOM	3718	CB	SER	A	470	14.849	39.747	10.216	1.00	44.26	C
ATOM	3719	OG	SER	A	470	15.979	39.120	10.800	1.00	50.97	O
ATOM	3720	N	ASP	A	471	13.844	42.835	11.209	1.00	44.35	N
ATOM	3721	CA	ASP	A	471	13.490	43.708	12.316	1.00	45.67	C
ATOM	3722	C	ASP	A	471	13.243	42.982	13.602	1.00	45.34	C
ATOM	3723	O	ASP	A	471	13.739	43.370	14.651	1.00	47.34	O
ATOM	3724	CB	ASP	A	471	12.244	44.503	11.962	1.00	47.32	C
ATOM	3725	CG	ASP	A	471	12.363	45.181	10.622	1.00	50.32	C
ATOM	3726	OD1	ASP	A	471	13.459	45.728	10.331	1.00	50.17	O
ATOM	3727	OD2	ASP	A	471	11.368	45.145	9.855	1.00	52.66	O
ATOM	3728	N	ARG	A	472	12.493	41.904	13.513	1.00	43.76	N
ATOM	3729	CA	ARG	A	472	12.148	41.155	14.691	1.00	43.10	C
ATOM	3730	C	ARG	A	472	13.409	40.620	15.400	1.00	41.30	C
ATOM	3731	O	ARG	A	472	13.468	40.629	16.643	1.00	40.99	O
ATOM	3732	CB	ARG	A	472	11.154	40.046	14.310	1.00	46.34	C
ATOM	3733	CG	ARG	A	472	10.306	40.396	13.056	1.00	48.84	C
ATOM	3734	CD	ARG	A	472	9.522	39.208	12.478	1.00	49.91	C
ATOM	3735	NE	ARG	A	472	8.353	38.880	13.288	1.00	51.36	N
ATOM	3736	CZ	ARG	A	472	7.843	37.661	13.416	1.00	51.68	C
ATOM	3737	NH1	ARG	A	472	8.403	36.637	12.792	1.00	52.78	N
ATOM	3738	NH2	ARG	A	472	6.769	37.471	14.173	1.00	53.44	N
ATOM	3739	N	VAL	A	473	14.433	40.235	14.626	1.00	37.44	N
ATOM	3740	CA	VAL	A	473	15.656	39.717	15.222	1.00	35.26	C
ATOM	3741	C	VAL	A	473	16.241	40.847	16.024	1.00	36.18	C
ATOM	3742	O	VAL	A	473	16.428	40.706	17.226	1.00	37.15	O
ATOM	3743	CB	VAL	A	473	16.690	39.219	14.212	1.00	33.44	C
ATOM	3744	CG1	VAL	A	473	17.986	38.923	14.919	1.00	33.16	C
ATOM	3745	CG2	VAL	A	473	16.233	37.960	13.569	1.00	31.26	C
ATOM	3746	N	THR	A	474	16.465	41.987	15.374	1.00	36.67	N
ATOM	3747	CA	THR	A	474	17.001	43.184	16.029	1.00	36.92	C
ATOM	3748	C	THR	A	474	16.129	43.593	17.240	1.00	37.16	C
ATOM	3749	O	THR	A	474	16.642	43.997	18.288	1.00	37.22	O
ATOM	3750	CB	THR	A	474	17.044	44.335	15.043	1.00	36.38	C
ATOM	3751	OG1	THR	A	474	17.757	43.925	13.883	1.00	38.97	O
ATOM	3752	CG2	THR	A	474	17.767	45.488	15.614	1.00	38.85	C
ATOM	3753	N	LYS	A	475	14.816	43.465	17.093	1.00	36.63	N
ATOM	3754	CA	LYS	A	475	13.887	43.785	18.164	1.00	36.72	C
ATOM	3755	C	LYS	A	475	14.194	42.904	19.392	1.00	35.59	C
ATOM	3756	O	LYS	A	475	14.606	43.426	20.434	1.00	34.54	O
ATOM	3757	CB	LYS	A	475	12.452	43.536	17.695	1.00	38.37	C
ATOM	3758	CG	LYS	A	475	11.415	43.847	18.736	1.00	40.43	C
ATOM	3759	CD	LYS	A	475	10.092	43.243	18.380	1.00	43.02	C
ATOM	3760	CE	LYS	A	475	9.189	43.265	19.601	1.00	46.14	C
ATOM	3761	NZ	LYS	A	475	8.166	42.159	19.508	1.00	49.81	N
ATOM	3762	N	CYS	A	476	13.982	41.584	19.267	1.00	33.09	N
ATOM	3763	CA	CYS	A	476	14.253	40.639	20.353	1.00	29.85	C
ATOM	3764	C	CYS	A	476	15.681	40.759	20.909	1.00	29.39	C
ATOM	3765	O	CYS	A	476	15.899	40.647	22.115	1.00	29.90	O
ATOM	3766	CB	CYS	A	476	14.056	39.210	19.882	1.00	29.33	C
ATOM	3767	SG	CYS	A	476	12.347	38.653	19.638	1.00	30.63	S
ATOM	3768	N	CYS	A	477	16.654	41.013	20.044	1.00	28.96	N
ATOM	3769	CA	CYS	A	477	18.047	41.110	20.480	1.00	29.96	C
ATOM	3770	C	CYS	A	477	18.395	42.358	21.260	1.00	31.95	C
ATOM	3771	O	CYS	A	477	19.079	42.283	22.284	1.00	32.58	O
ATOM	3772	CB	CYS	A	477	19.022	40.975	19.294	1.00	27.52	C
ATOM	3773	SG	CYS	A	477	19.122	39.349	18.480	1.00	26.69	S
ATOM	3774	N	THR	A	478	17.976	43.505	20.723	1.00	34.42	N
ATOM	3775	CA	THR	A	478	18.213	44.838	21.289	1.00	35.00	C
ATOM	3776	C	THR	A	478	17.343	45.238	22.476	1.00	35.97	C
ATOM	3777	O	THR	A	478	17.806	45.903	23.373	1.00	38.11	O
ATOM	3778	CB	THR	A	478	18.025	45.908	20.182	1.00	34.79	C
ATOM	3779	OG1	THR	A	478	19.094	45.814	19.228	1.00	34.17	O

FIG. 12D

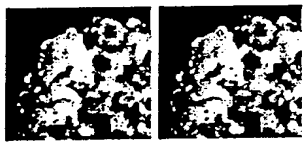
ATOM	9140	CE	LYS	B	573	63.370	-4.593	43.427	1.00	107.11	C
ATOM	9141	NZ	LYS	B	573	63.010	-5.996	43.088	1.00	106.97	N
ATOM	9142	N	LYS	B	574	59.690	0.439	44.459	1.00	102.70	N
ATOM	9143	CA	LYS	B	574	59.315	1.053	45.726	1.00	102.18	C
ATOM	9144	C	LYS	B	574	57.823	0.930	46.006	1.00	100.98	C
ATOM	9145	O	LYS	B	574	57.400	0.838	47.162	1.00	100.83	O
ATOM	9146	CB	LYS	B	574	59.743	2.533	45.760	1.00	103.36	C
ATOM	9147	CG	LYS	B	574	61.279	2.769	45.749	1.00	104.79	C
ATOM	9148	CD	LYS	B	574	61.992	2.237	47.025	1.00	105.34	C
ATOM	9149	CE	LYS	B	574	63.518	2.095	46.833	1.00	105.48	C
ATOM	9150	NZ	LYS	B	574	64.229	1.471	48.005	1.00	105.59	N
ATOM	9151	N	LEU	B	575	57.029	0.878	44.943	1.00	100.22	N
ATOM	9152	CA	LEU	B	575	55.577	0.789	45.082	1.00	99.84	C
ATOM	9153	C	LEU	B	575	54.989	-0.605	45.318	1.00	98.71	C
ATOM	9154	O	LEU	B	575	53.860	-0.738	45.818	1.00	98.12	O
ATOM	9155	CB	LEU	B	575	54.884	1.520	43.927	1.00	100.66	C
ATOM	9156	CG	LEU	B	575	55.329	2.994	43.843	1.00	101.40	C
ATOM	9157	CD1	LEU	B	575	54.423	3.797	42.918	1.00	102.02	C
ATOM	9158	CD2	LEU	B	575	55.352	3.629	45.232	1.00	101.72	C
ATOM	9159	N	VAL	B	576	55.758	-1.636	44.980	1.00	97.58	N
ATOM	9160	CA	VAL	B	576	55.321	-3.007	45.210	1.00	96.95	C
ATOM	9161	C	VAL	B	576	55.567	-3.347	46.679	1.00	96.77	C
ATOM	9162	O	VAL	B	576	54.974	-4.278	47.230	1.00	97.43	O
ATOM	9163	CB	VAL	B	576	56.072	-4.018	44.324	1.00	96.69	C
ATOM	9164	CG1	VAL	B	576	55.678	-3.833	42.863	1.00	97.29	C
ATOM	9165	CG2	VAL	B	576	57.566	-3.869	44.508	1.00	96.25	C
ATOM	9166	N	ALA	B	577	56.476	-2.604	47.300	1.00	96.22	N
ATOM	9167	CA	ALA	B	577	56.782	-2.787	48.711	1.00	95.43	C
ATOM	9168	C	ALA	B	577	55.657	-2.143	49.527	1.00	94.78	C
ATOM	9169	O	ALA	B	577	55.246	-2.672	50.563	1.00	94.54	O
ATOM	9170	CB	ALA	B	577	58.112	-2.130	49.041	1.00	95.29	C
ATOM	9171	N	ALA	B	578	55.139	-1.030	49.005	1.00	94.36	N
ATOM	9172	CA	ALA	B	578	54.063	-0.251	49.628	1.00	94.60	C
ATOM	9173	C	ALA	B	578	52.695	-0.916	49.624	1.00	95.09	C
ATOM	9174	O	ALA	B	578	51.904	-0.732	50.555	1.00	95.20	O
ATOM	9175	CB	ALA	B	578	53.958	1.102	48.965	1.00	94.06	C
ATOM	9176	N	SER	B	579	52.385	-1.607	48.531	1.00	95.84	N
ATOM	9177	CA	SER	B	579	51.116	-2.316	48.395	1.00	96.35	C
ATOM	9178	C	SER	B	579	51.124	-3.633	49.195	1.00	96.58	C
ATOM	9179	O	SER	B	579	50.093	-4.055	49.718	1.00	96.30	O
ATOM	9180	CB	SER	B	579	50.823	-2.569	46.906	1.00	96.57	C
ATOM	9181	OG	SER	B	579	51.993	-3.003	46.219	1.00	96.64	O
ATOM	9182	N	GLN	B	580	52.301	-4.258	49.301	1.00	96.86	N
ATOM	9183	CA	GLN	B	580	52.493	-5.515	50.042	1.00	97.04	C
ATOM	9184	C	GLN	B	580	52.105	-5.310	51.511	1.00	96.75	C
ATOM	9185	O	GLN	B	580	51.554	-6.208	52.159	1.00	96.88	O
ATOM	9186	CB	GLN	B	580	53.968	-5.944	49.947	1.00	97.83	C
ATOM	9187	CG	GLN	B	580	54.282	-7.403	50.343	1.00	99.77	C
ATOM	9188	CD	GLN	B	580	54.215	-7.663	51.852	1.00	101.01	C
ATOM	9189	OE1	GLN	B	580	53.842	-8.755	52.291	1.00	101.84	O
ATOM	9190	NE2	GLN	B	580	54.586	-6.659	52.648	1.00	101.42	N
ATOM	9191	N	ALA	B	581	52.476	-4.148	52.047	1.00	96.39	N
ATOM	9192	CA	ALA	B	581	52.166	-3.791	53.429	1.00	95.65	C
ATOM	9193	C	ALA	B	581	50.665	-3.676	53.689	1.00	94.79	C
ATOM	9194	O	ALA	B	581	50.201	-3.980	54.790	1.00	94.77	O
ATOM	9195	CB	ALA	B	581	52.879	-2.493	53.806	1.00	95.87	C
ATOM	9196	N	ALA	B	582	49.920	-3.199	52.695	1.00	93.61	N
ATOM	9197	CA	ALA	B	582	48.469	-3.080	52.819	1.00	92.86	C
ATOM	9198	C	ALA	B	582	47.848	-4.479	52.984	1.00	92.50	C
ATOM	9199	O	ALA	B	582	46.874	-4.625	53.757	1.00	91.96	O
ATOM	9200	CB	ALA	B	582	47.907	-2.377	51.591	1.00	92.58	C
ATOM	9201	OXT	ALA	B	582	48.371	-5.426	52.357	1.00	91.45	O
TER	9202		ALA	B	582						
HETATM	9203	O	HOH		601	35.839	-17.824	20.755	1.00	15.91	O
HETATM	9204	O	HOH		602	38.450	-27.051	23.636	1.00	38.88	O
HETATM	9205	O	HOH		603	35.109	4.954	17.718	1.00	47.18	O
HETATM	9206	O	HOH		604	23.897	-18.840	30.789	1.00	19.48	O

FIG. 12E

```
HETATM 9207  O   HOH   605      26.814  33.912  22.514  1.00 18.60           O
HETATM 9208  O   HOH   606      30.418  23.669  30.769  1.00 40.43           O
HETATM 9209  O   HOH   607      28.287  20.400  14.901  1.00 36.69           O
CONECT  398  397  461
CONECT  461  398  460
CONECT  563  562  682
CONECT  676  675  764
CONECT  682  563  681
CONECT  764  676  763
CONECT  954  953 1347
CONECT 1341 1340 1399
CONECT 1347  954 1346
CONECT 1399 1341 1398
CONECT 1579 1578 1944
CONECT 1938 1937 1997
CONECT 1944 1579 1943
CONECT 1997 1938 1996
CONECT 2090 2089 2197
CONECT 2191 2190 2278
CONECT 2197 2090 2196
CONECT 2278 2191 2277
CONECT 2479 2478 2855
CONECT 2849 2848 2910
CONECT 2855 2479 2854
CONECT 2910 2849 2909
CONECT 3103 3102 3467
CONECT 3461 3460 3548
CONECT 3467 3103 3466
CONECT 3548 3461 3547
CONECT 3649 3648 3773
CONECT 3767 3766 3853
CONECT 3773 3649 3772
CONECT 3853 3767 3852
CONECT 4073 4072 4432
CONECT 4426 4425 4493
CONECT 4432 4073 4431
CONECT 4493 4426 4492
CONECT 4999 4998 5062
CONECT 5062 4999 5061
CONECT 5164 5163 5283
CONECT 5277 5276 5365
CONECT 5283 5164 5282
CONECT 5365 5277 5364
CONECT 5555 5554 5948
CONECT 5942 5941 6000
CONECT 5948 5555 5947
CONECT 6000 5942 5999
CONECT 6180 6179 6545
CONECT 6539 6538 6598
CONECT 6545 6180 6544
CONECT 6598 6539 6597
CONECT 6691 6690 6798
CONECT 6792 6791 6879
CONECT 6798 6691 6797
CONECT 6879 6792 6878
CONECT 7080 7079 7456
CONECT 7450 7449 7511
CONECT 7456 7080 7455
CONECT 7511 7450 7510
CONECT 7704 7703 8068
CONECT 8062 8061 8149
CONECT 8068 7704 8067
CONECT 8149 8062 8148
CONECT 8250 8249 8374
CONECT 8368 8367 8454
CONECT 8374 8250 8373
CONECT 8454 8368 8453
```

FIG. 12F

SBdBase



Visualization Toolkit

For further information or assistance,
contact dbmaster@strubix.com

Interactive visualization of the protein.

FIG. 13

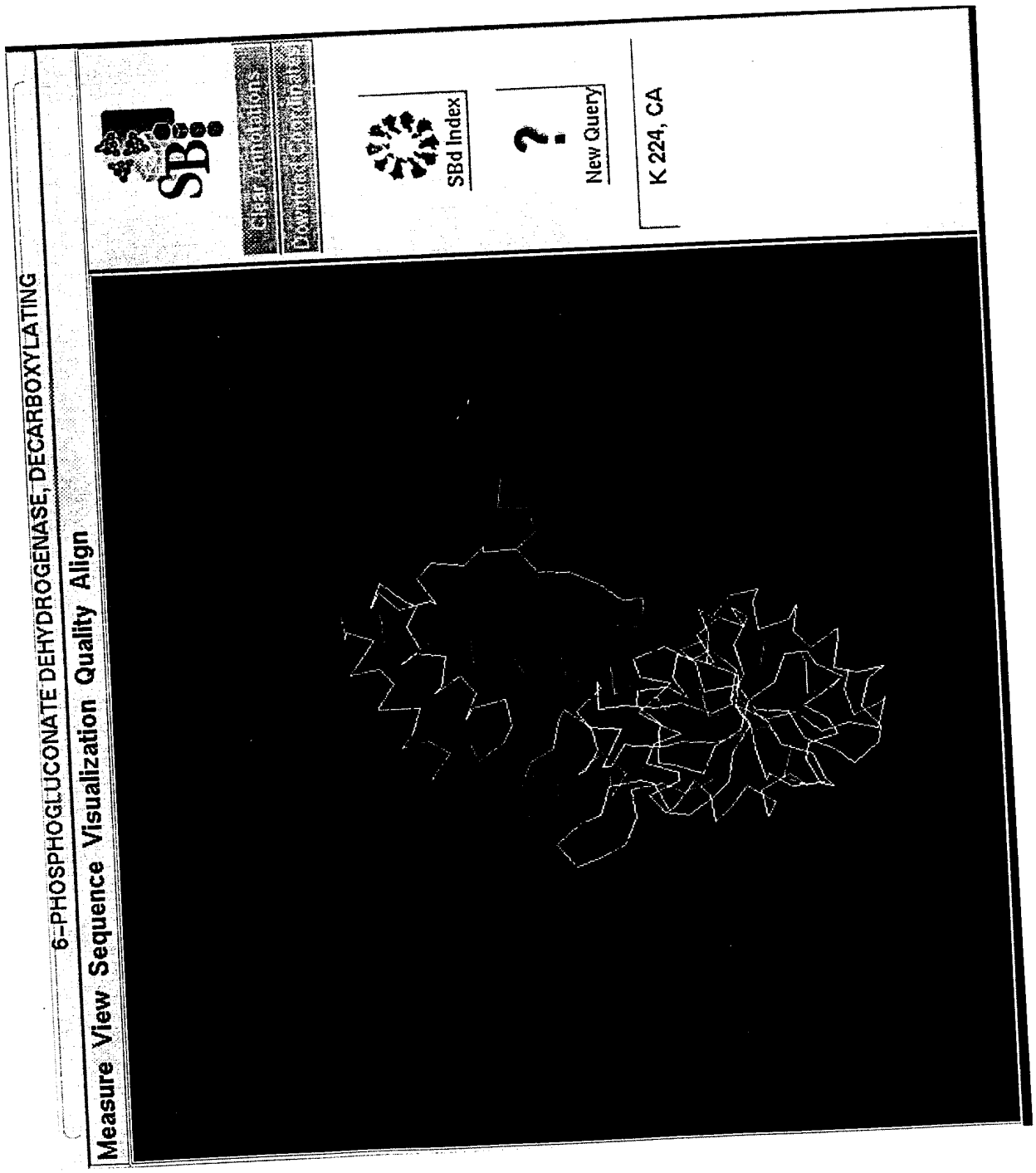

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SBd Index
New Query
K 224, CA

FIG. 14

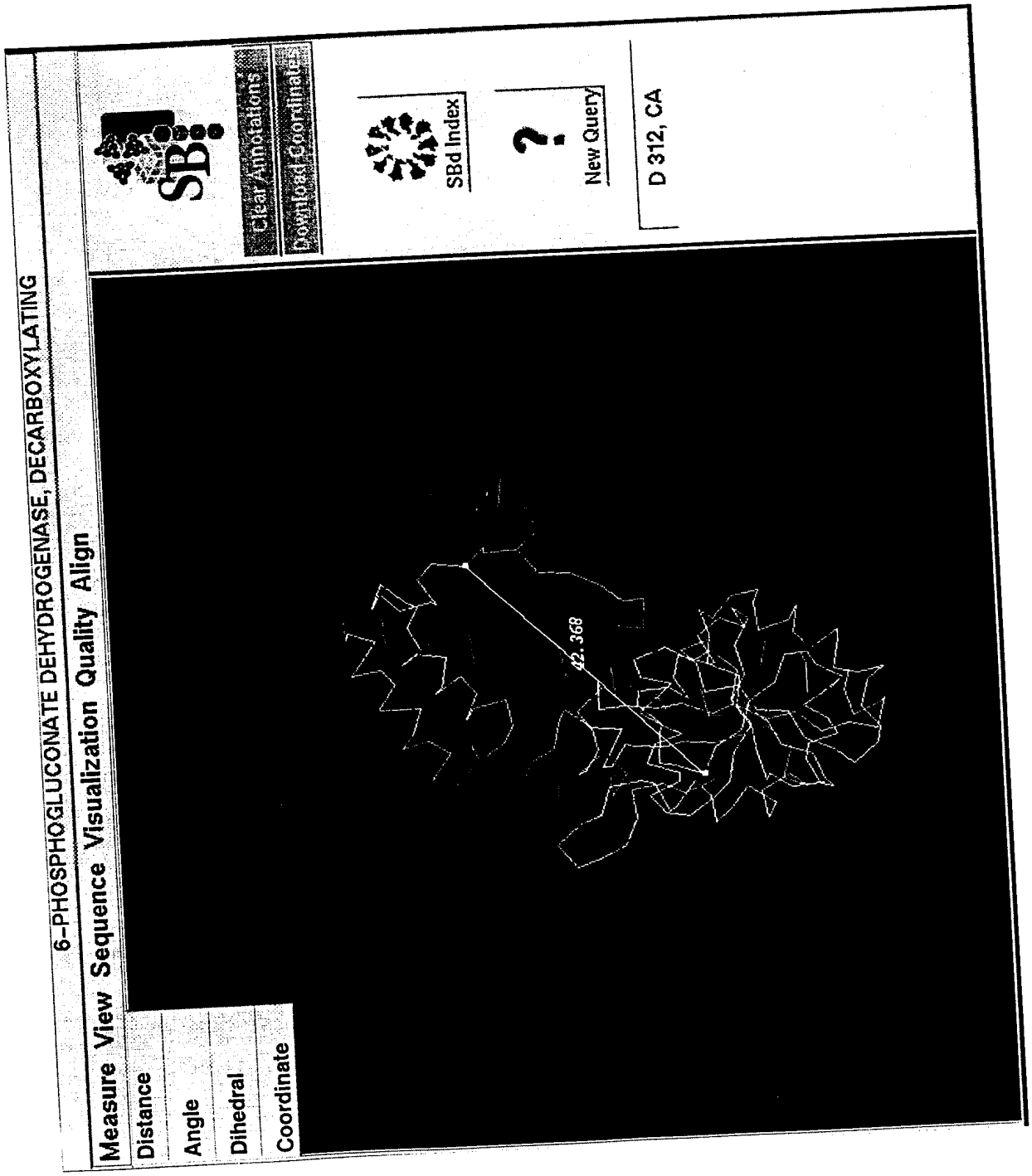

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
Distance
Angle
Dihedral
Coordinate
SBd
Clear Annotations
Download Coordinates
SBd Index
?
New Query
D 312, CA
42.368

FIG. 15

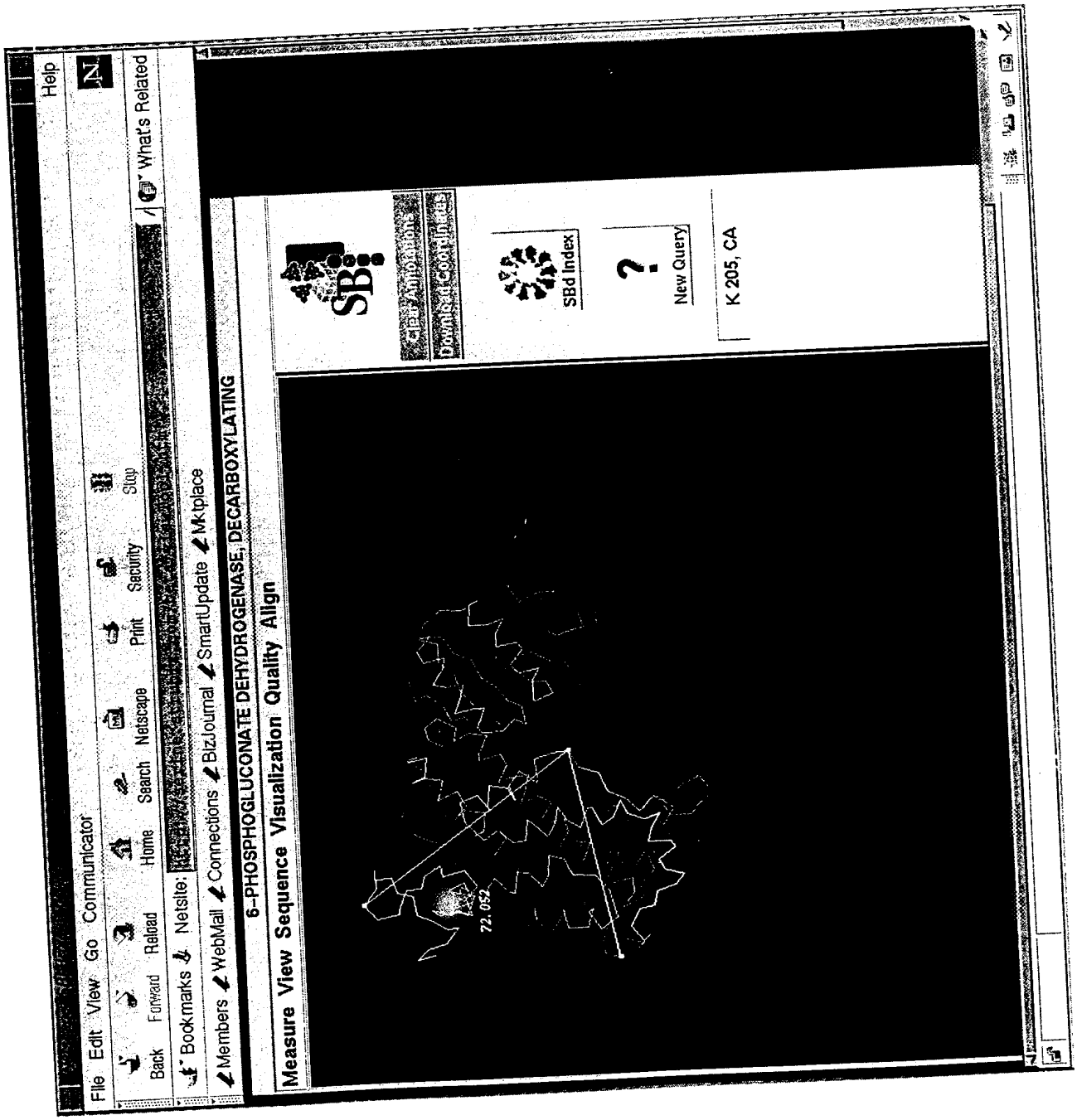

File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Netsite:
Members WebMail Connections BizJournal SmartUpdate Mktplace
6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SBd Index
New Query
K 205, CA

FIG. 16

Netscape: SBI Protein Viewer
File Edit View Go Communicator
Help
Back Reload Home Search Netscape Print Security What's Related
Bookmarks Netsite: http://serine.strubix.com/beta/viewTool.htm
Members WebMail Connections BizJournal SmartUpdate Mktplace
6-PHOSPHOGLUCONATE DEHYDROGENASE. DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SBd Index
New Query
G 472, CA

FIG. 17

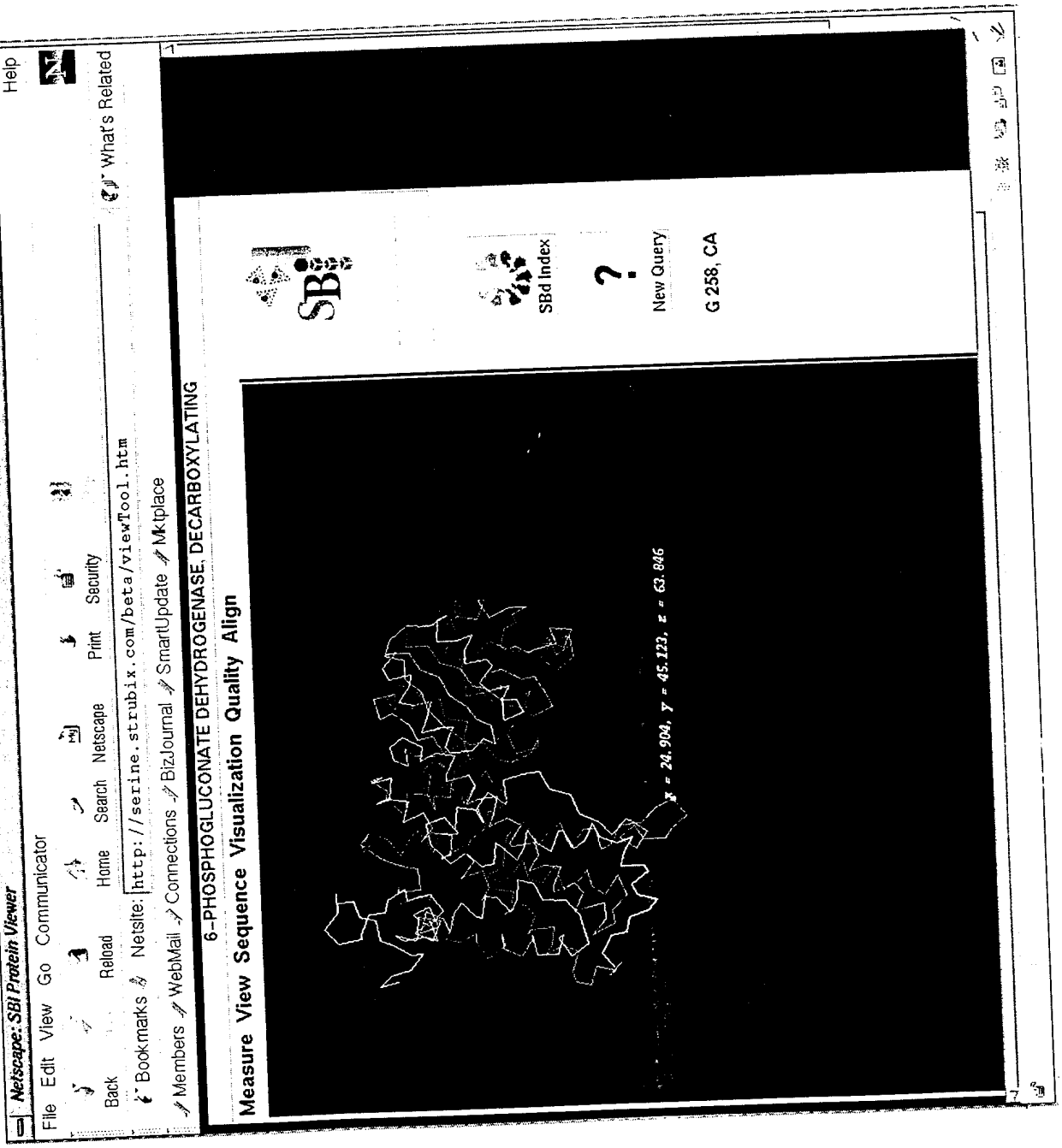

Netscape: SBI Protein Viewer
File Edit View Go Communicator
Help
Back Reload Home Search Netscape Print Security
What's Related
Bookmarks Netsite: http://serine.strubix.com/beta/viewTool.htm
Members WebMail Connections BizJournal SmartUpdate Mktplace
6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
x = 24.904, y = 45.123, z = 63.846
SBd Index
New Query
G 258, CA

FIG. 18

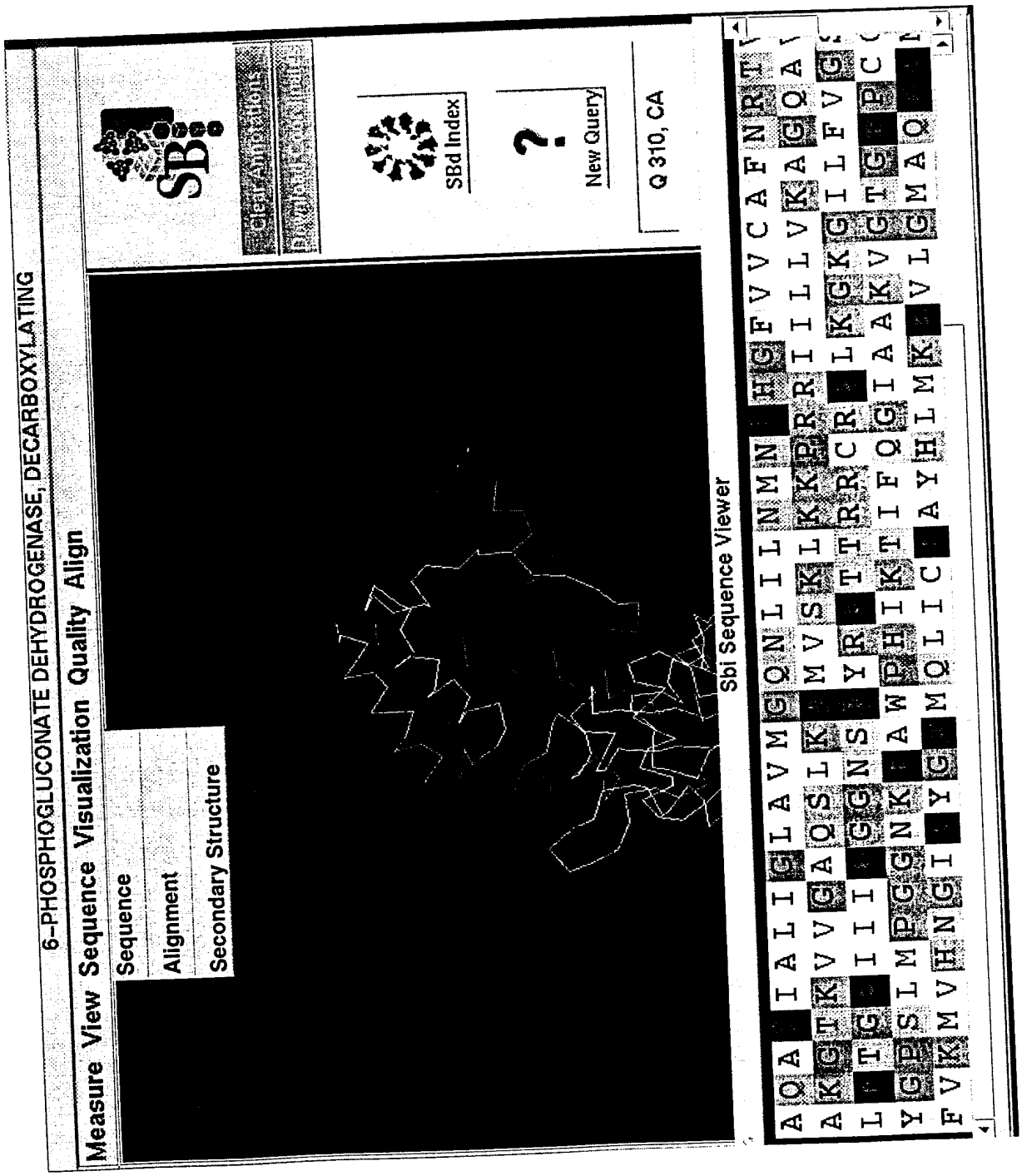

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
Sequence
Alignment
Secondary Structure
SBd Index
New Query
Q 310, CA
Sbi Sequence Viewer

FIG. 19

Netscape: SBI Protein Viewer
File Edit View Go Communicator
Help
Back Reload Home Search Netscape Print Security
Bookmarks Netsite: http://seriine.strubix.com/beta/viewTool.htm
What's Related
Members WebMail Connections BizJournal SmartUpdate Mktplace
6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SBd Index
New Query
D 245, CA
Sequence Alignment
6-PG1-DH 1 AQA IALIGLAVMGQNLILNMN HGFVVCAFNRTVSKV FLAN AKGTKVVGAQSLK MVSKL
2PGD 1 AQA IALIGLAVMGQNLILNMN HGFVVCAFNRTVSKV FLAN AKGTKVLGAHSL MVSKL
6-PG1-DH 76 AGQAV FI KLVPLL TG III GGNS YR TTRRCR LKGKGILFVGSGVSGG GFRYGPS
Unsigned Java Applet Window

FIG. 20

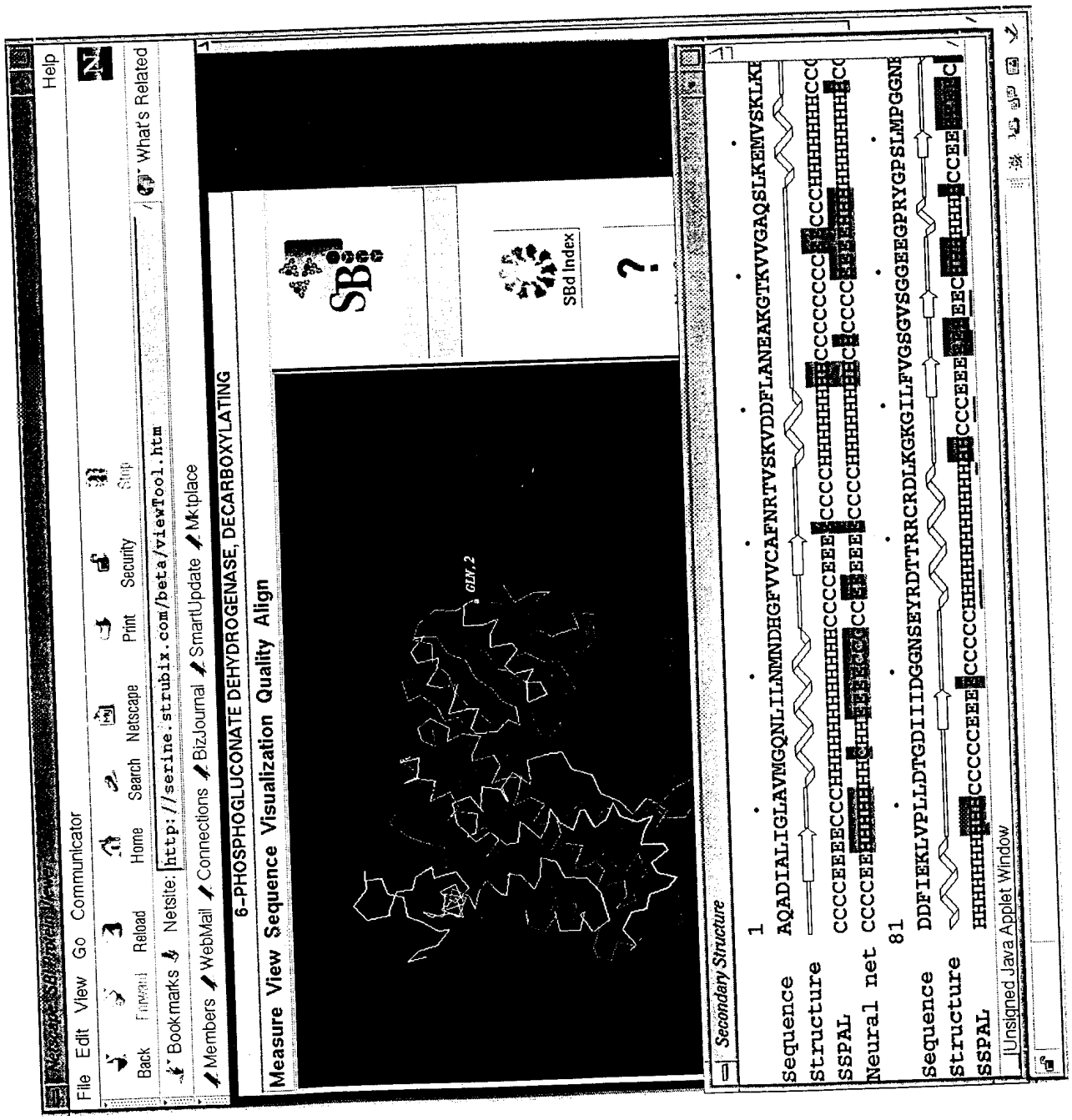

File Edit View Go Communicator
Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Netsite: http://seriine.strubix.com/beta/viewTool.htm
What's Related
Members WebMail Connections BizJournal SmartUpdate Mktplace
6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SBd Index
Secondary Structure
Sequence
Structure
SSPAL
Neural net
Sequence
Structure
SSPAL
Unsigned Java Applet Window

FIG. 21

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
Hydrophobicity
Active Sites
Glycosylation
Phosphorylation
Natural Variants
Secondary Structure Ribbon
Electrostatic Surface
Dynamic Surface
SBd Index
New Query
K 316, CA

FIG. 22

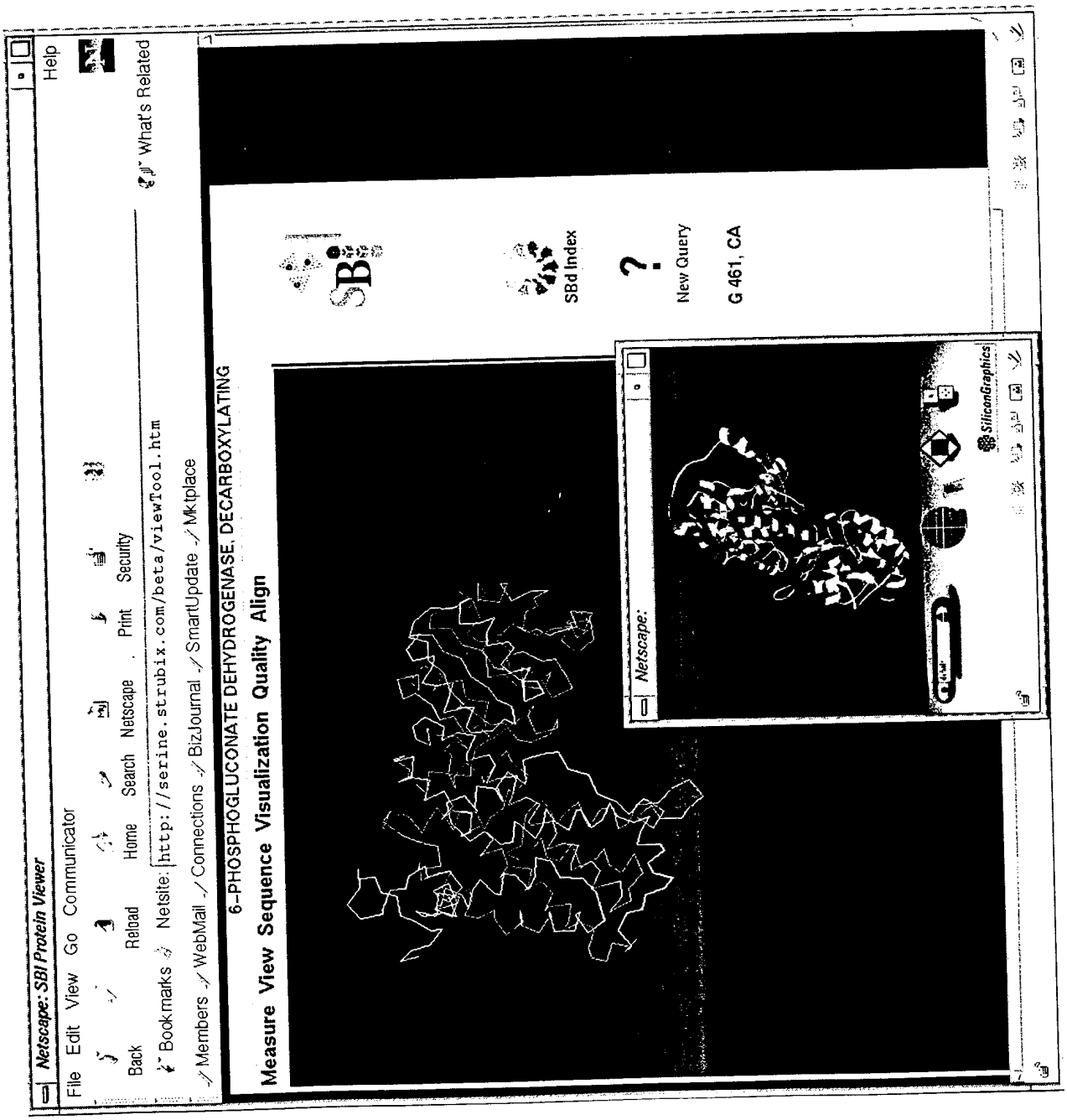

Netscape: SBI Protein Viewer
File Edit View Go Communicator
Help
Back Reload Home Search Netscape Print Security
Bookmarks Netsite: http://serine.strubix.com/beta/viewTool.htm
What's Related
Members WebMail Connections BizJournal SmartUpdate Mktplace
6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SBd Index
New Query
G 461, CA
Netscape:
SiliconGraphics

FIG. 23

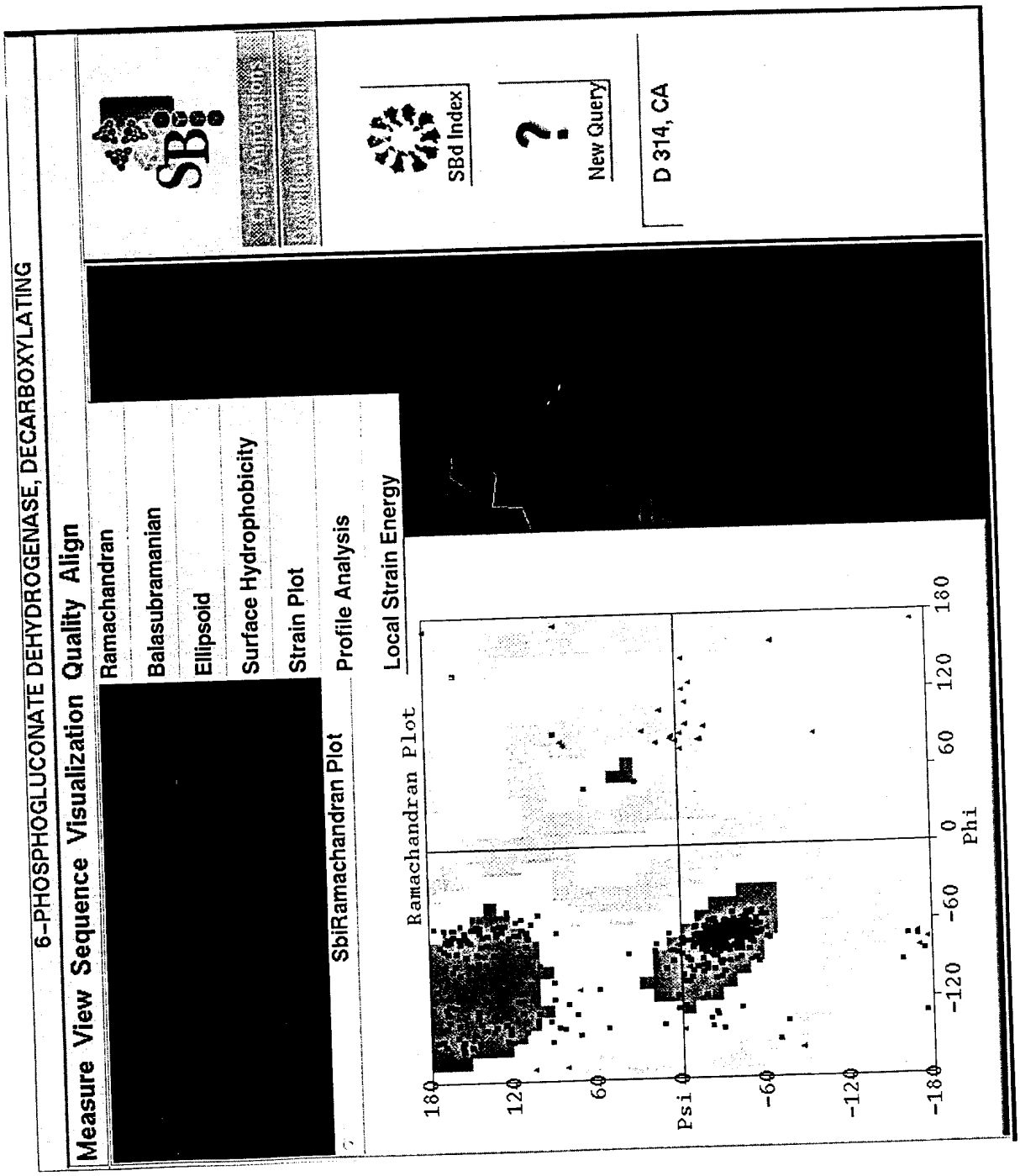

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
Ramachandran
Balasubramanian
Ellipsoid
Surface Hydrophobicity
Strain Plot
SbiRamachandran Plot
Profile Analysis
Local Strain Energy
SBd Index
New Query
D 314, CA
Ramachandran Plot
Psi
Phi
180
120
60
0
-60
-120
-180

FIG. 24

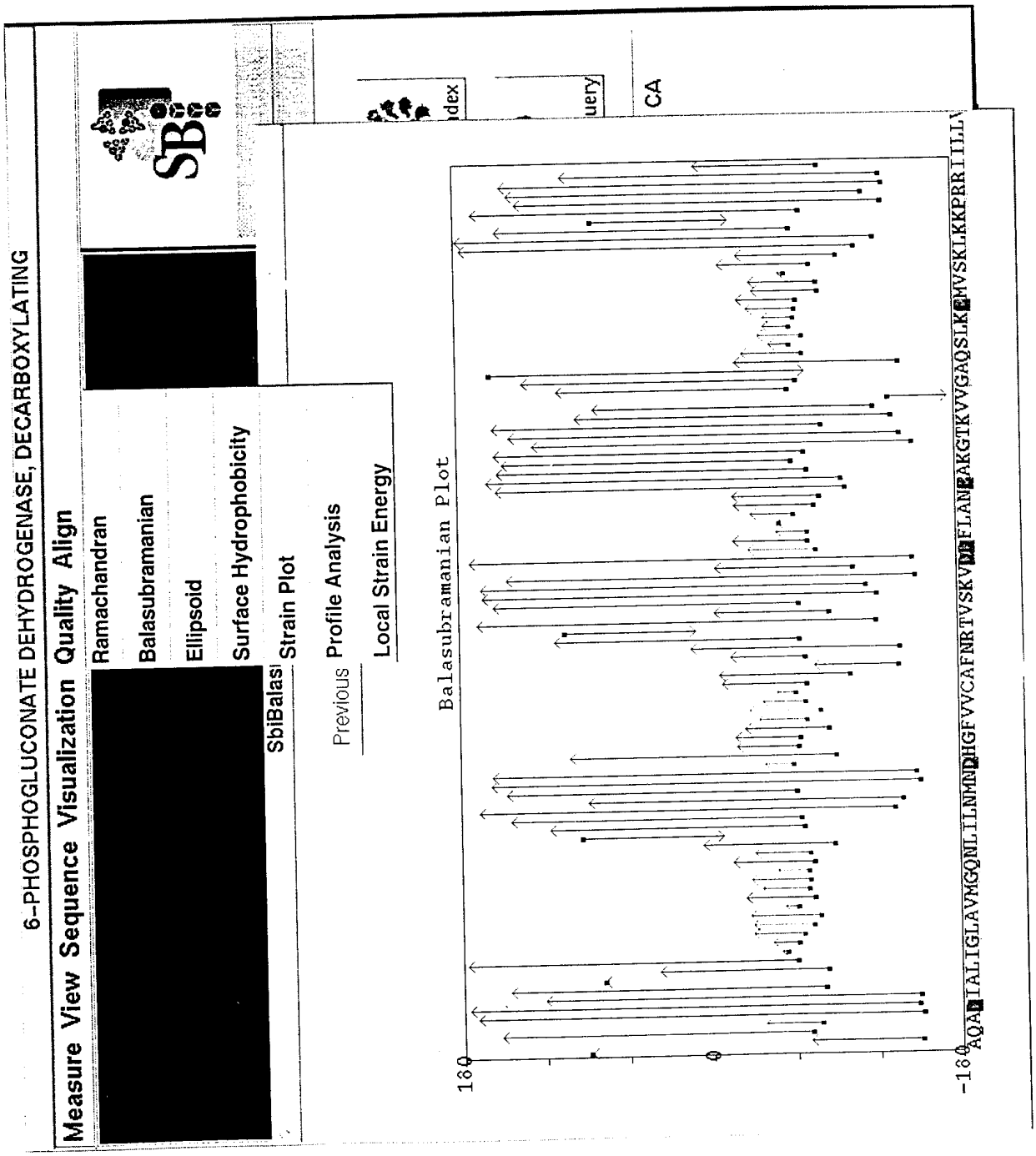

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
Ramachandran
Balasubramanian
Ellipsoid
Surface Hydrophobicity
SbiBalasi Strain Plot
Previous Profile Analysis
Local Strain Energy
Balasubramanian Plot
CA
100
-100

FIG. 25

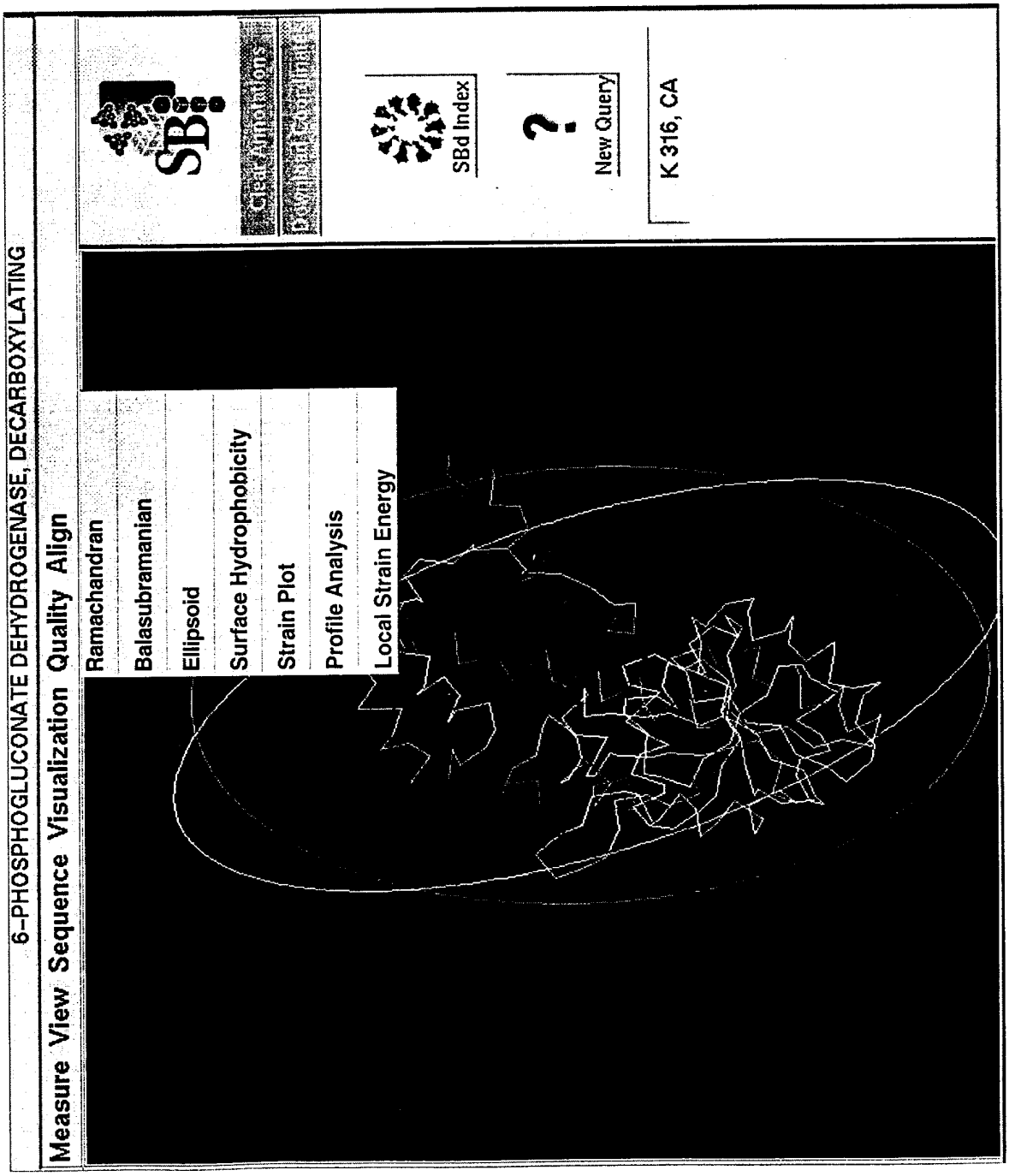

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
Ramachandran
Balasubramanian
Ellipsoid
Surface Hydrophobicity
Strain Plot
Profile Analysis
Local Strain Energy
SBd Index
New Query
K 316, CA

FIG. 26

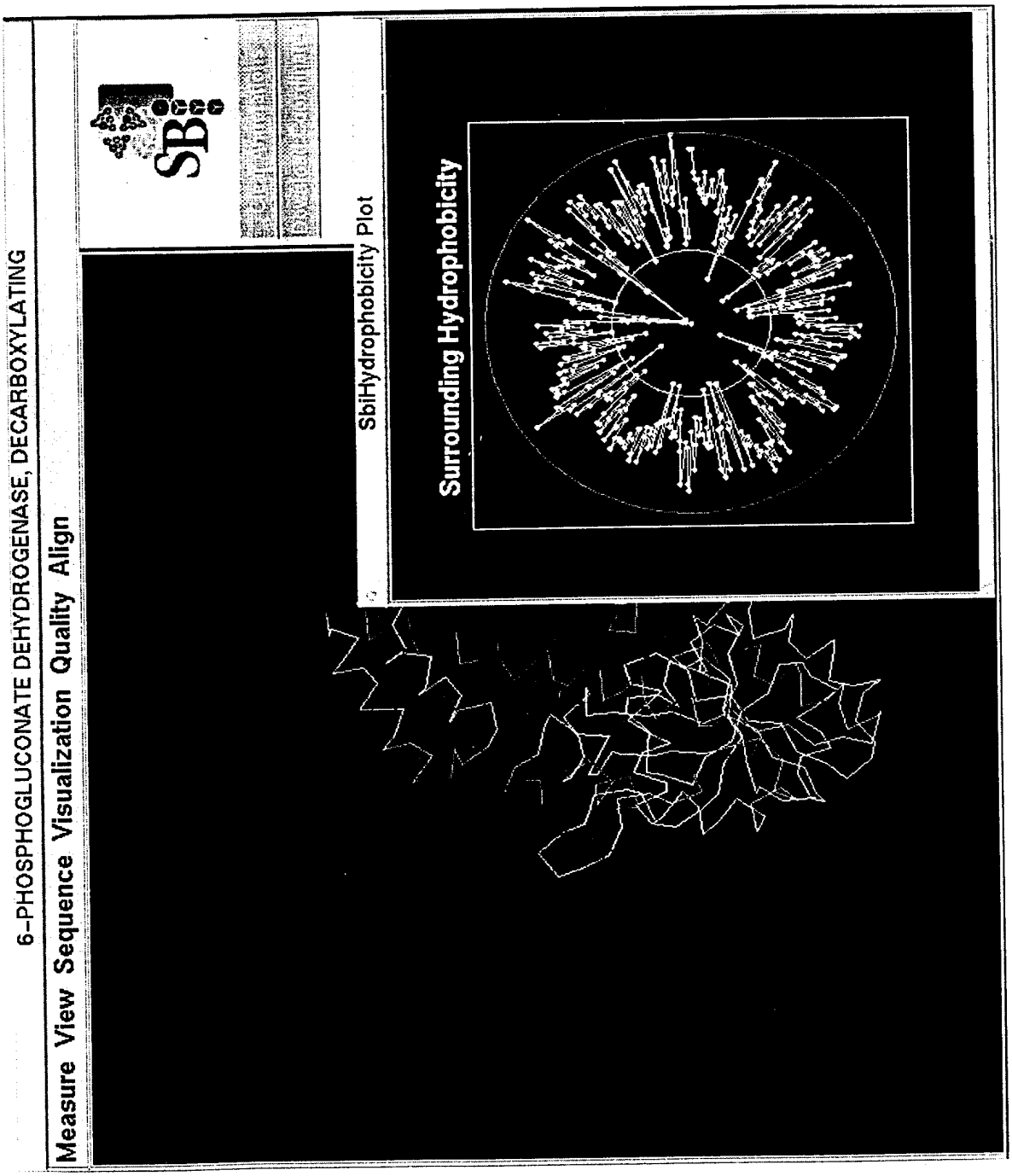

6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SBi
SbiHydrophobicity Plot
Surrounding Hydrophobicity

FIG. 27

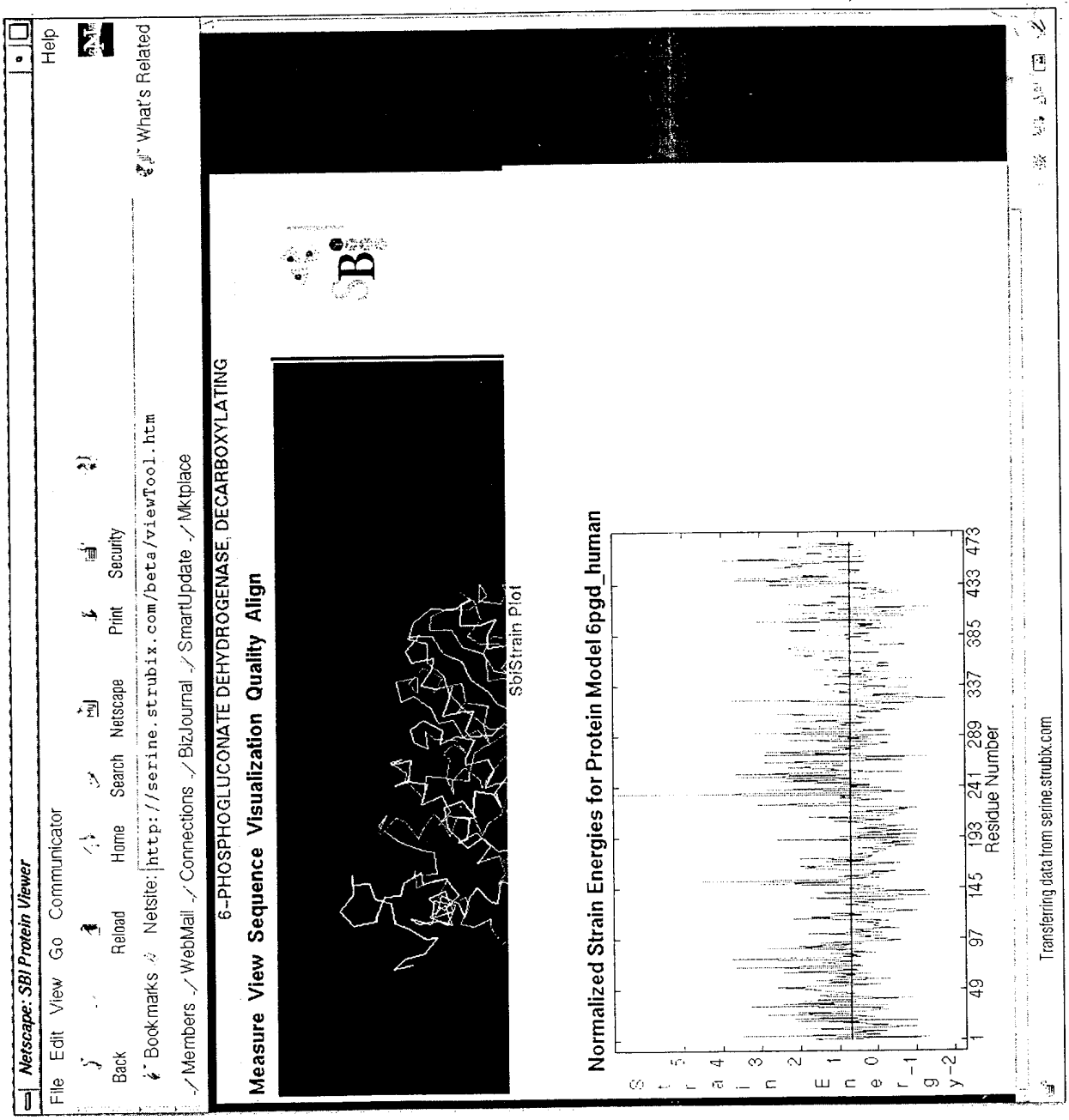

Netscape: SBI Protein Viewer
File Edit View Go Communicator
Help
Back Reload Home Search Netscape Print Security
Bookmarks Netsite: http://serine.strubix.com/beta/viewTool.htm
What's Related
Members WebMail Connections BizJournal SmartUpdate Mktplace
6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SbiStrain Plot
Normalized Strain Energies for Protein Model 6pgd_human
Strain Energy
5
4
3
2
1
0
-1
-2
1 49 97 145 193 241 289 337 385 433 473
Residue Number
Transferring data from serine.strubix.com

FIG. 28

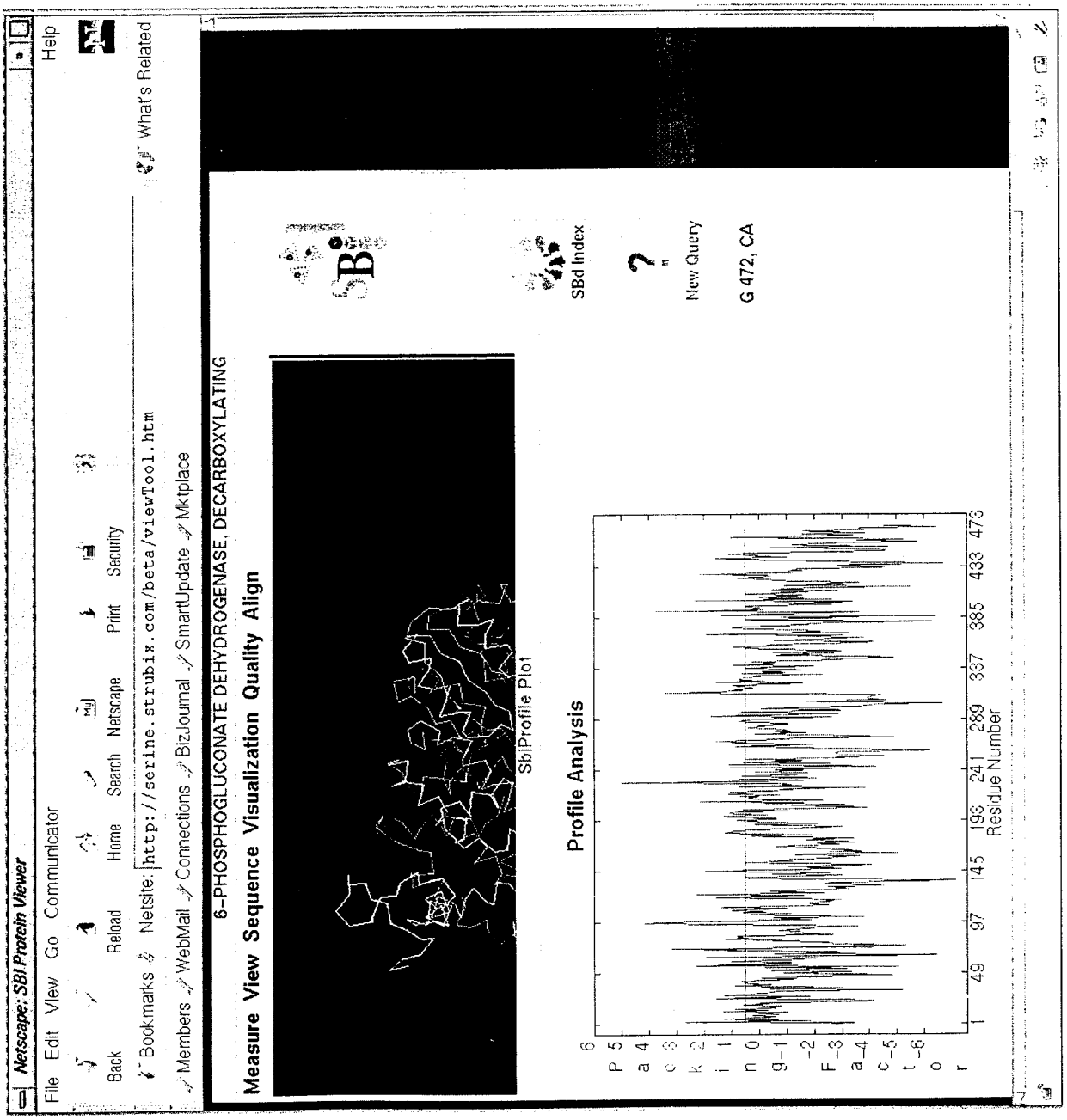

Netscape: SBI Protein Viewer
File Edit View Go Communicator Help
Back Reload Home Search Netscape Print Security
Bookmarks Netsite: http://serine.strubix.com/beta/viewTool.htm What's Related
Members WebMail Connections BizJournal SmartUpdate Mktplace
6-PHOSPHOGLUCONATE DEHYDROGENASE, DECARBOXYLATING
Measure View Sequence Visualization Quality Align
SbiProfile Plot
SBd Index
New Query
G 472, CA
Profile Analysis
Residue Number
1 49 97 145 196 241 289 337 385 433 473

FIG. 29

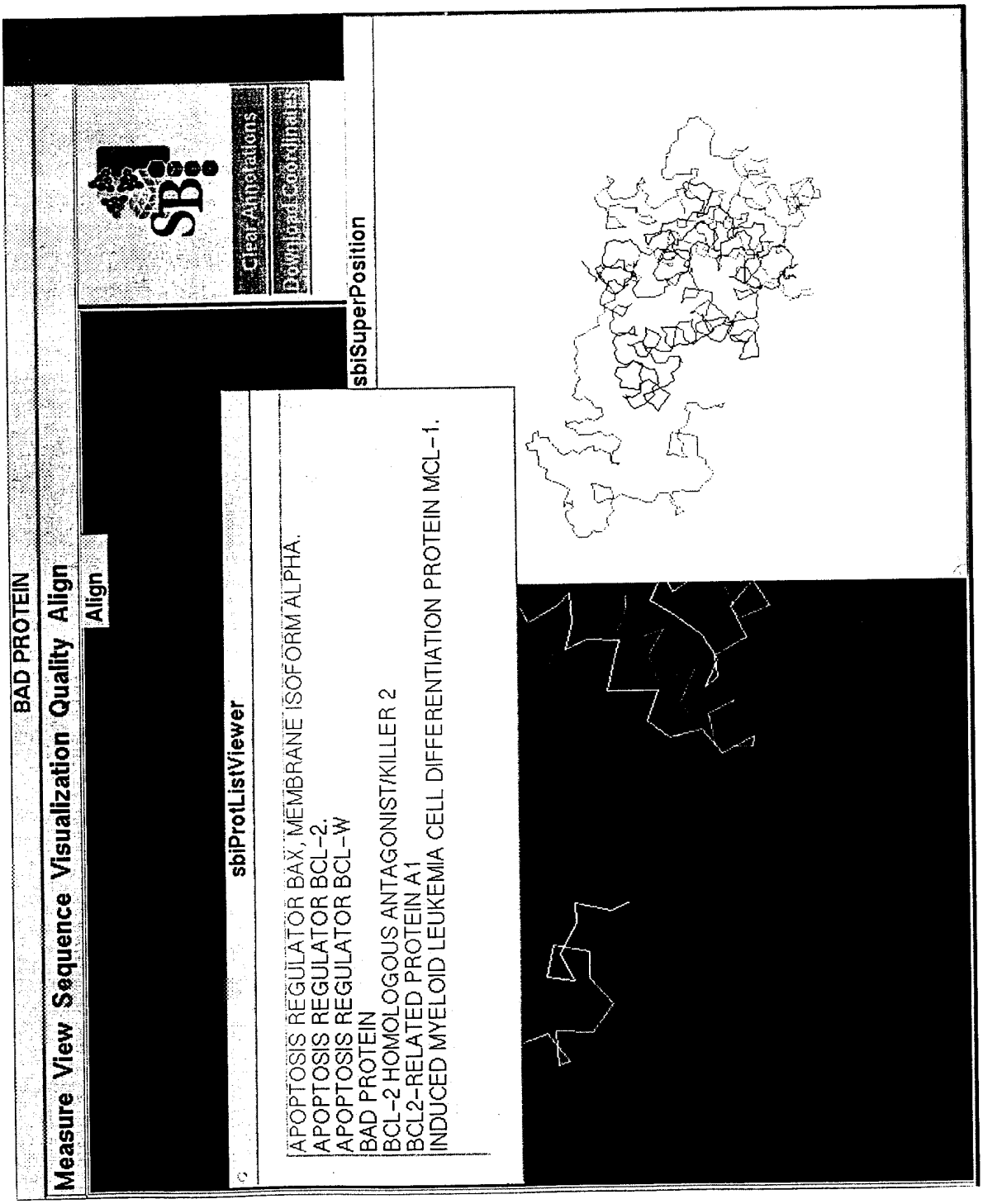

BAD PROTEIN
Measure View Sequence Visualization Quality Align
Align
sbiProtListViewer
APOPTOSIS REGULATOR BAX, MEMBRANE ISOFORM ALPHA.
APOPTOSIS REGULATOR BCL-2.
APOPTOSIS REGULATOR BCL-W
BAD PROTEIN
BCL-2 HOMOLOGOUS ANTAGONIST/KILLER 2
BCL2-RELATED PROTEIN A1
INDUCED MYELOID LEUKEMIA CELL DIFFERENTIATION PROTEIN MCL-1.
SBI
Clear Annotations
Download Coordinates
sbiSuperPosition

FIG. 30

SBdBase Protein Entry Form

Fields marked (*) are required

Family: * 6pg_dehydrogenase New *(enter name here)*
C-lectin
actin-depolymerizing

Protein name: *

Species: *

Annotation file:

Alignment:

Secondary structure:

Ribbon setup (ribbon):

Ribbon data (ribinline):

Natural variant coords:

Electrostatics surface:

Surface hydrophobicity data:

Profile analysis data:

Protein ellipsoid data:

Accessibility data:

Local strain data:

Local strain ribbon:

Local strain ribinline:

Dynamic surface:

Initial model:

- Initial coordinates:
- Ramachandran plot:
- Bond length graphs:
- Bond angle graphs:
- Depositor:

Refinement level 1:

- Level 1 coordinates:
- Ramachandran plot:
- Bond length graphs:
- Bond angle graphs:
- Depositor:

Refinement level 2:

FIG. 31A

Level 2 coordinates:
Ramachandran plot:
Bond length graphs:
Bond angle graphs:
Depositor:

Submit Clear

$Revision: 3.1 $

FIG. 31B

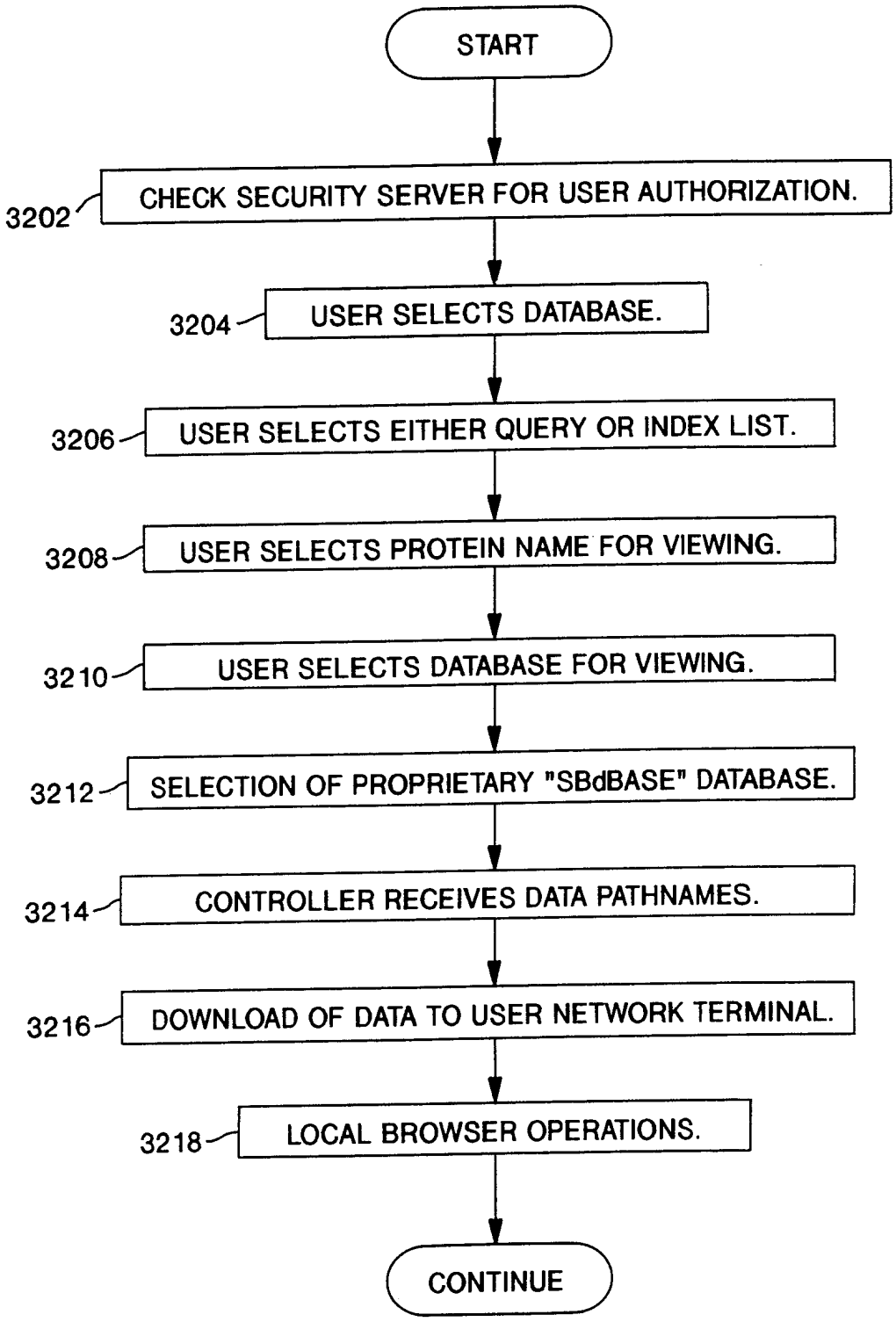

START
3202
CHECK SECURITY SERVER FOR USER AUTHORIZATION.
3204
USER SELECTS DATABASE.
3206
USER SELECTS EITHER QUERY OR INDEX LIST.
3208
USER SELECTS PROTEIN NAME FOR VIEWING.
3210
USER SELECTS DATABASE FOR VIEWING.
3212
SELECTION OF PROPRIETARY "SBdBASE" DATABASE.
3214
CONTROLLER RECEIVES DATA PATHNAMES.
3216
DOWNLOAD OF DATA TO USER NETWORK TERMINAL.
3218
LOCAL BROWSER OPERATIONS.
CONTINUE

FIG. 32

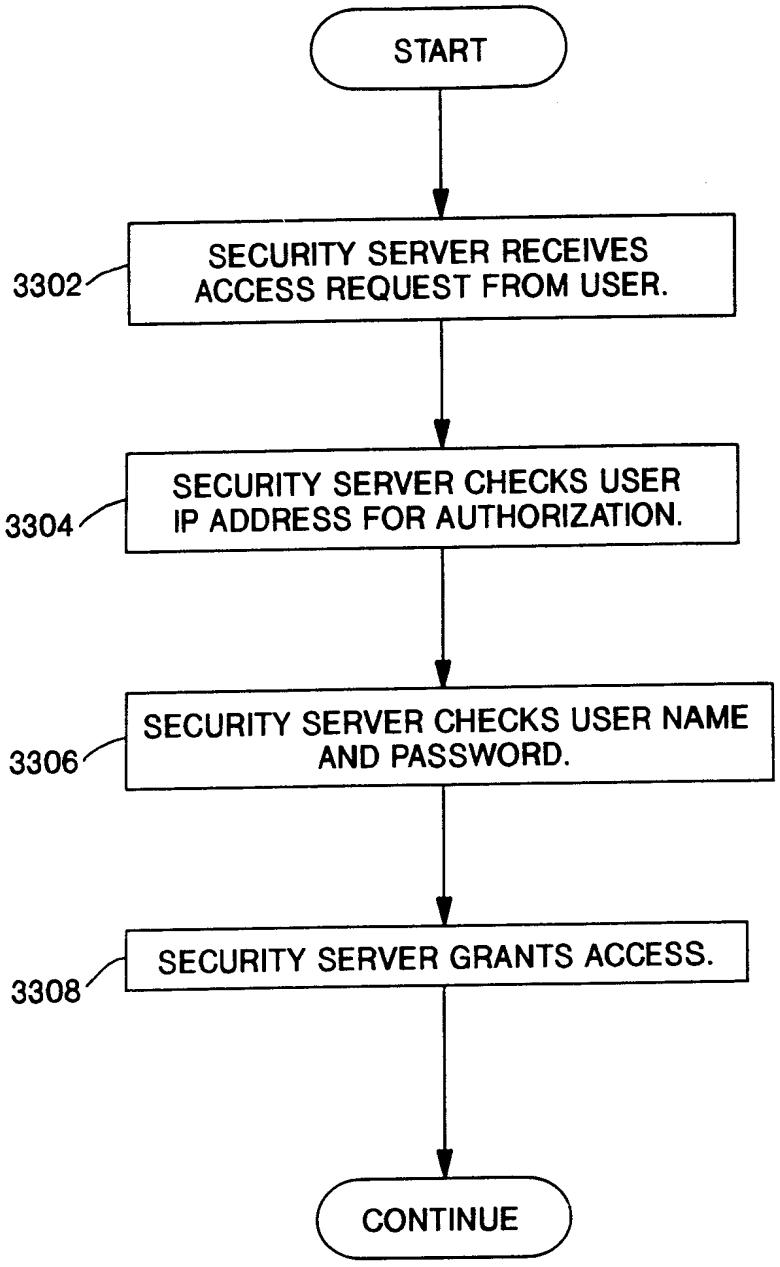

START
SECURITY SERVER RECEIVES ACCESS REQUEST FROM USER.
3302
SECURITY SERVER CHECKS USER IP ADDRESS FOR AUTHORIZATION.
3304
SECURITY SERVER CHECKS USER NAME AND PASSWORD.
3306
SECURITY SERVER GRANTS ACCESS.
3308
CONTINUE

FIG. 33

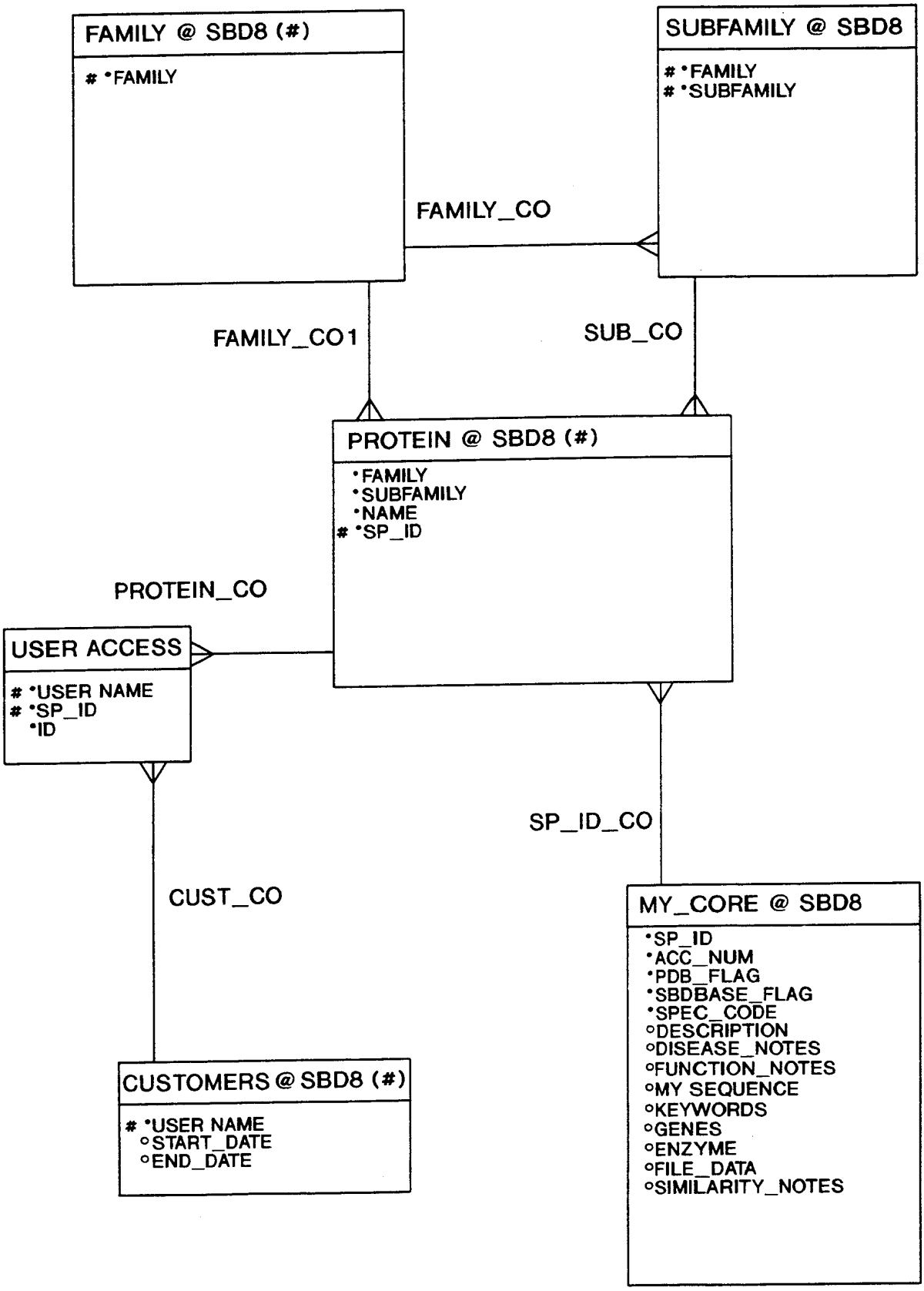

FAMILY @ SBD8 (#)
*FAMILY
SUBFAMILY @ SBD8
*FAMILY
*SUBFAMILY
FAMILY_CO
FAMILY_CO1
SUB_CO
PROTEIN @ SBD8 (#)
*FAMILY
*SUBFAMILY
*NAME
*SP_ID
PROTEIN_CO
USER ACCESS
*USER NAME
*SP_ID
*ID
SP_ID_CO
CUST_CO
MY_CORE @ SBD8
*SP_ID
*ACC_NUM
*PDB_FLAG
*SBDBASE_FLAG
*SPEC_CODE
oDESCRIPTION
oDISEASE_NOTES
oFUNCTION_NOTES
oMY SEQUENCE
oKEYWORDS
oGENES
oENZYME
oFILE_DATA
oSIMILARITY_NOTES
CUSTOMERS @ SBD8 (#)
*USER NAME
oSTART_DATE
oEND_DATE

FIG. 34

- class netscape.application.View (implements netscape.util.Codable)
  - class Ellipsoid
  - class HydrophobicityPlot
  - class netscape.application.InternalWindow (implements netscape.application.Window)
    - class sbiBaluCanvas
    - class sbiHydrophobicity
    - class sbiList
    - class sbiProfile
    - class sbiProtViewer
    - class sbiRama (implements netscape.application.WindowOwner)
    - class sbiSeqViewer
    - class sbiStrain
  - class ProfilePlot
  - class StrainPlot
  - class graphTest
  - class hydrophobicityGraph
  - class profileGraph
  - class sbiAlignViewer (implements netscape.application.Target)
  - class sbiBalu
  - class sbiBaluButtons (implements netscape.application.Target)
  - class sbiEllipsoid
  - class sbiGui (implements netscape.application.Target)
  - class sbiPdbCanvas (implements netscape.application.Target)
  - class sbiPdbViewer (implements netscape.application.Target)
  - class sbiSlider (implements netscape.application.Target)
- class sbiActiveSites
- class sbiConverter

FIG. 35

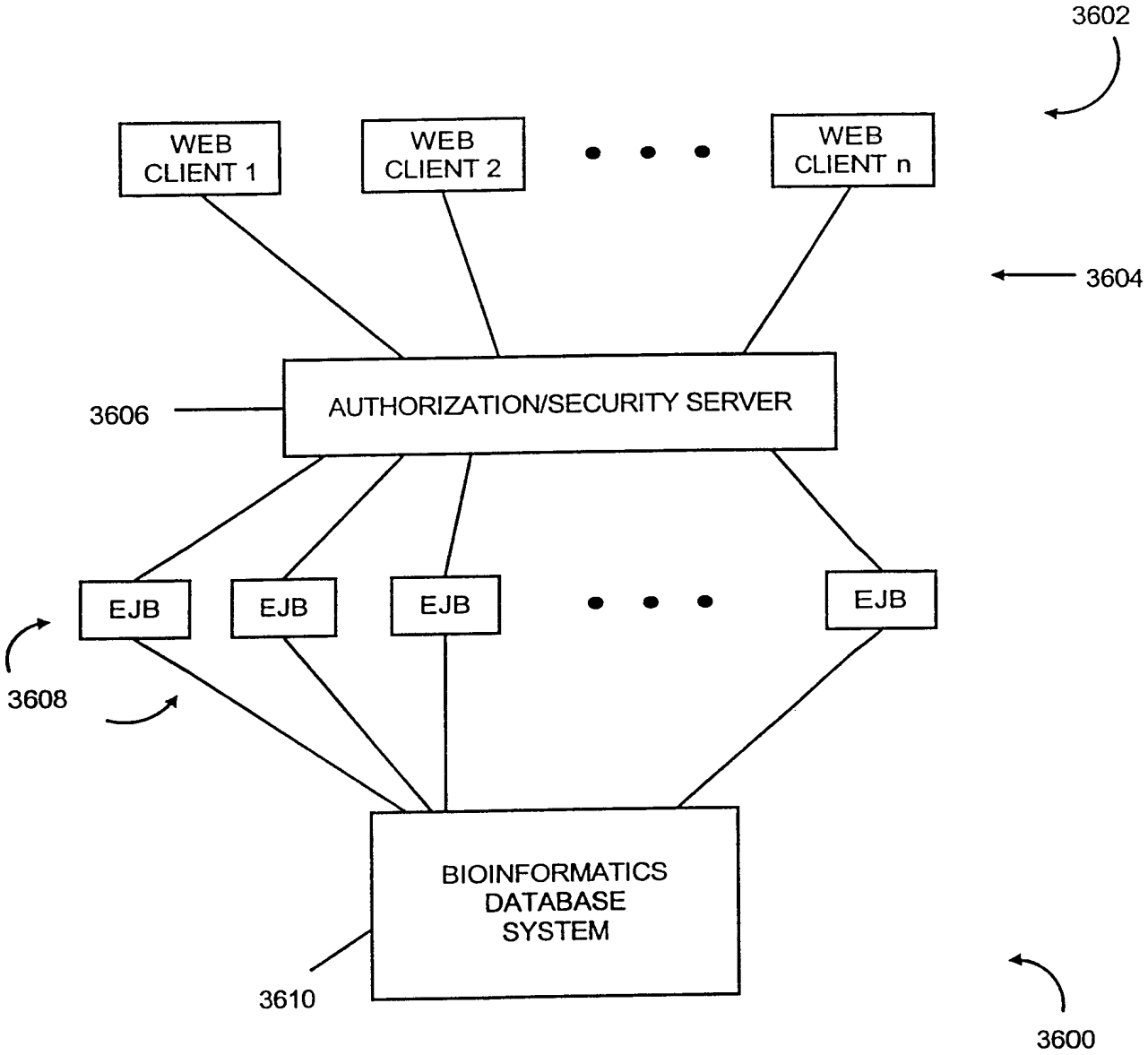

3602
WEB CLIENT 1
WEB CLIENT 2
WEB CLIENT n
3604
3606
AUTHORIZATION/SECURITY SERVER
EJB
EJB
EJB
EJB
3608
BIOINFORMATICS DATABASE SYSTEM
3610
3600

FIG. 36

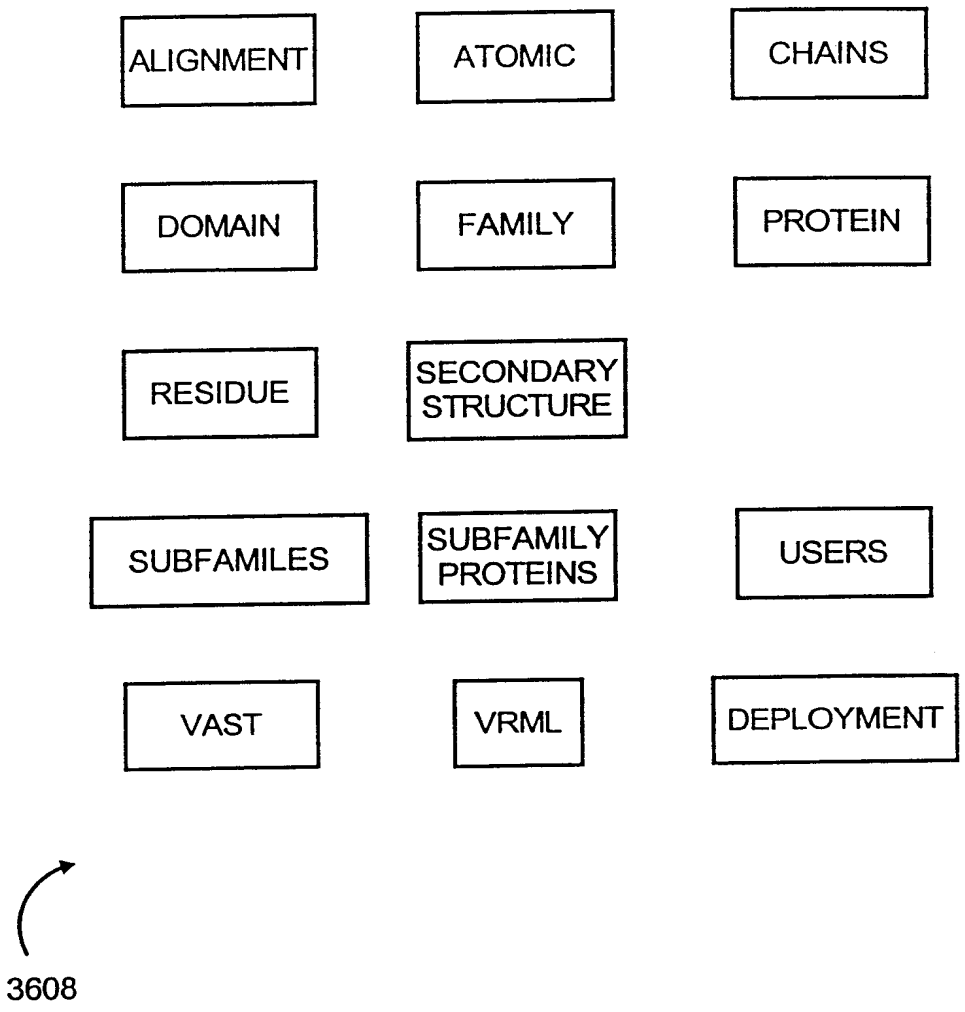

ALIGNMENT
ATOMIC
CHAINS
DOMAIN
FAMILY
PROTEIN
RESIDUE
SECONDARY STRUCTURE
SUBFAMILES
SUBFAMILY PROTEINS
USERS
VAST
VRML
DEPLOYMENT
3608

FIG. 37

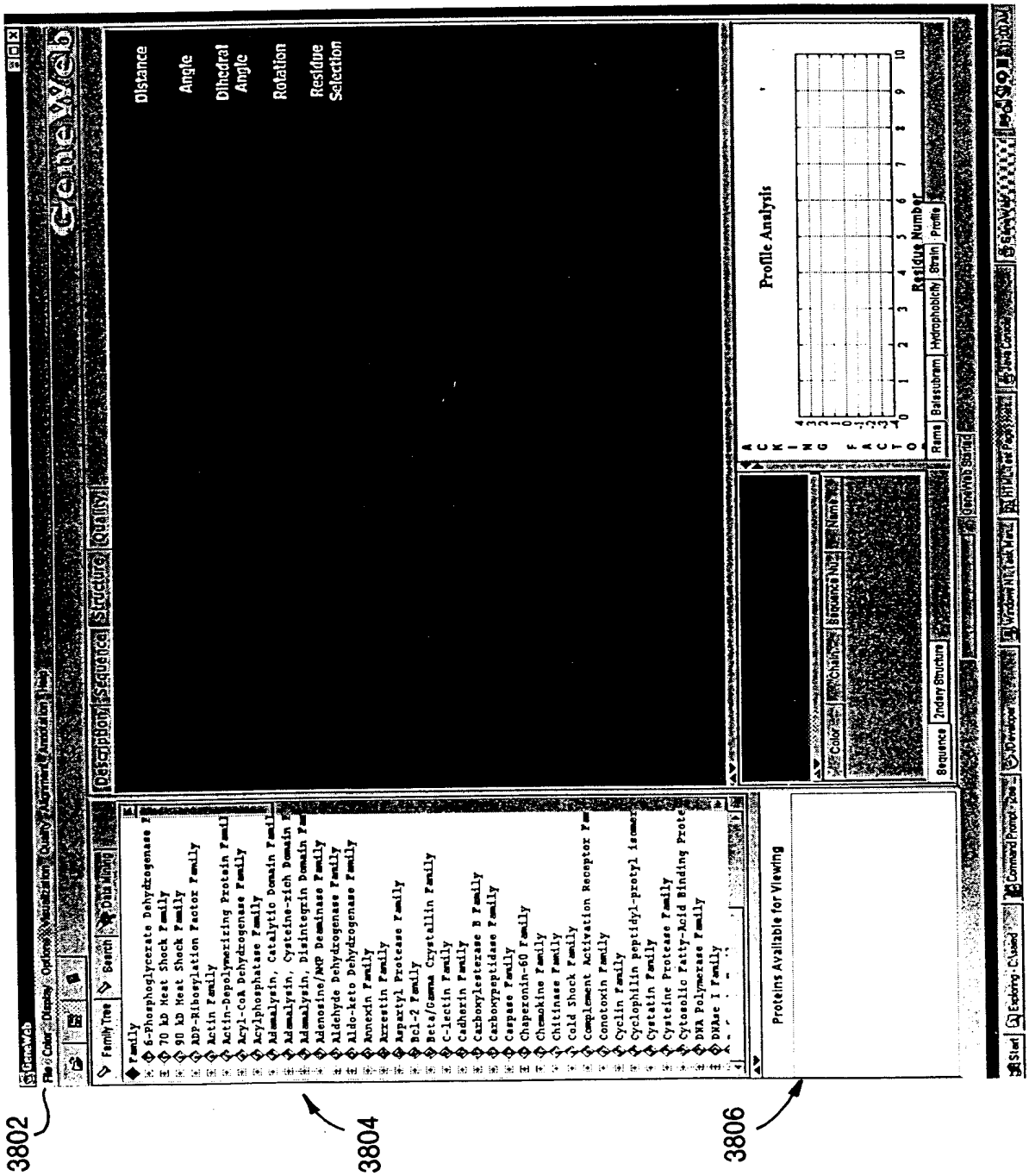

3800
3802
3804
3806
3808
3810
GeneWeb
Distance
Angle
Dihedral Angle
Rotation
Residue Selection
Profile Analysis
Residue Number
Proteins Available for Viewing

FIG. 38

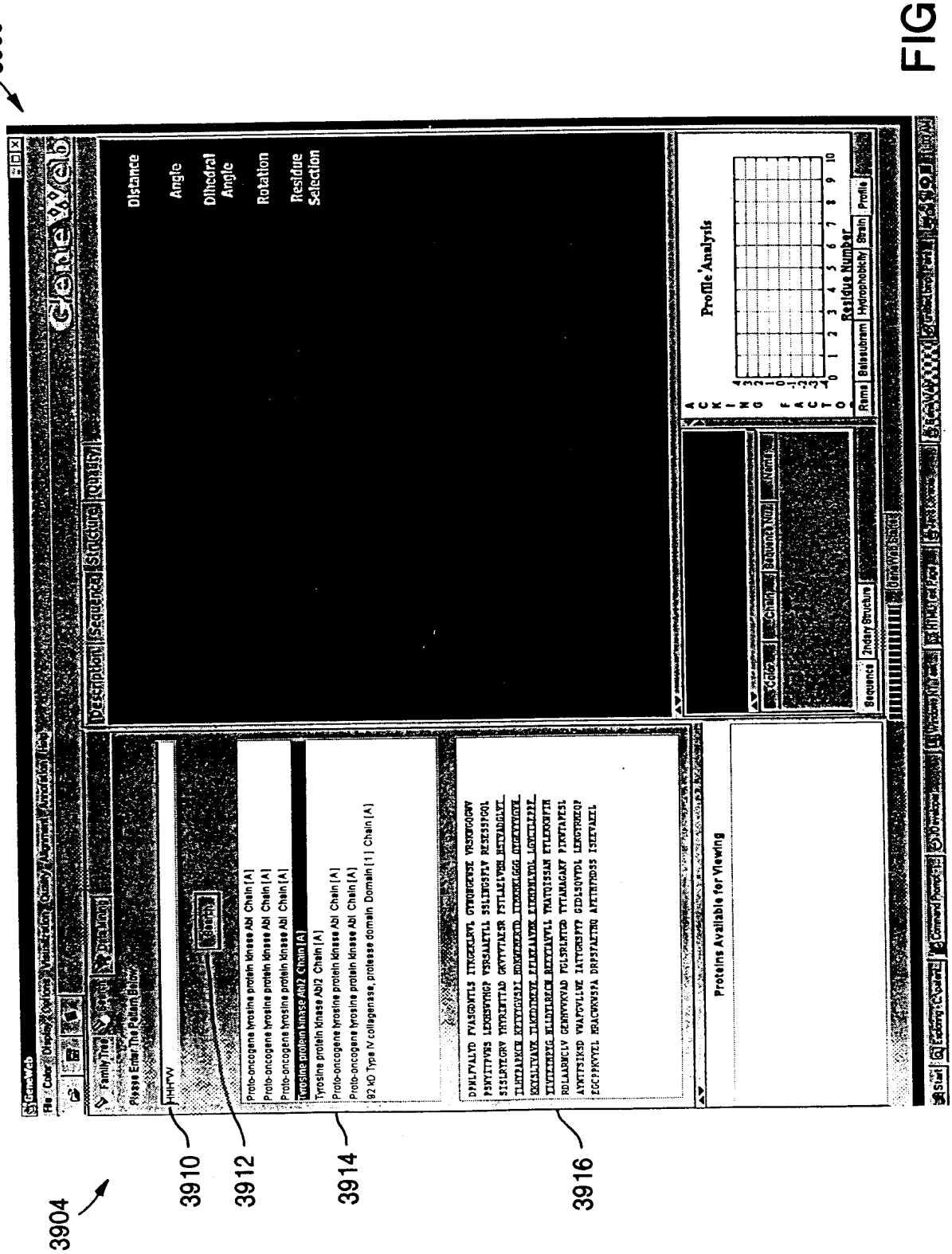

3900
3904
3910
3912
3914
3916
GeneWeb
Description
Sequence
Structure
Distance
Angle
Dihedral Angle
Rotation
Residue Selection
Profile Analysis
Residue Number
Proteins Available for Viewing

FIG. 39

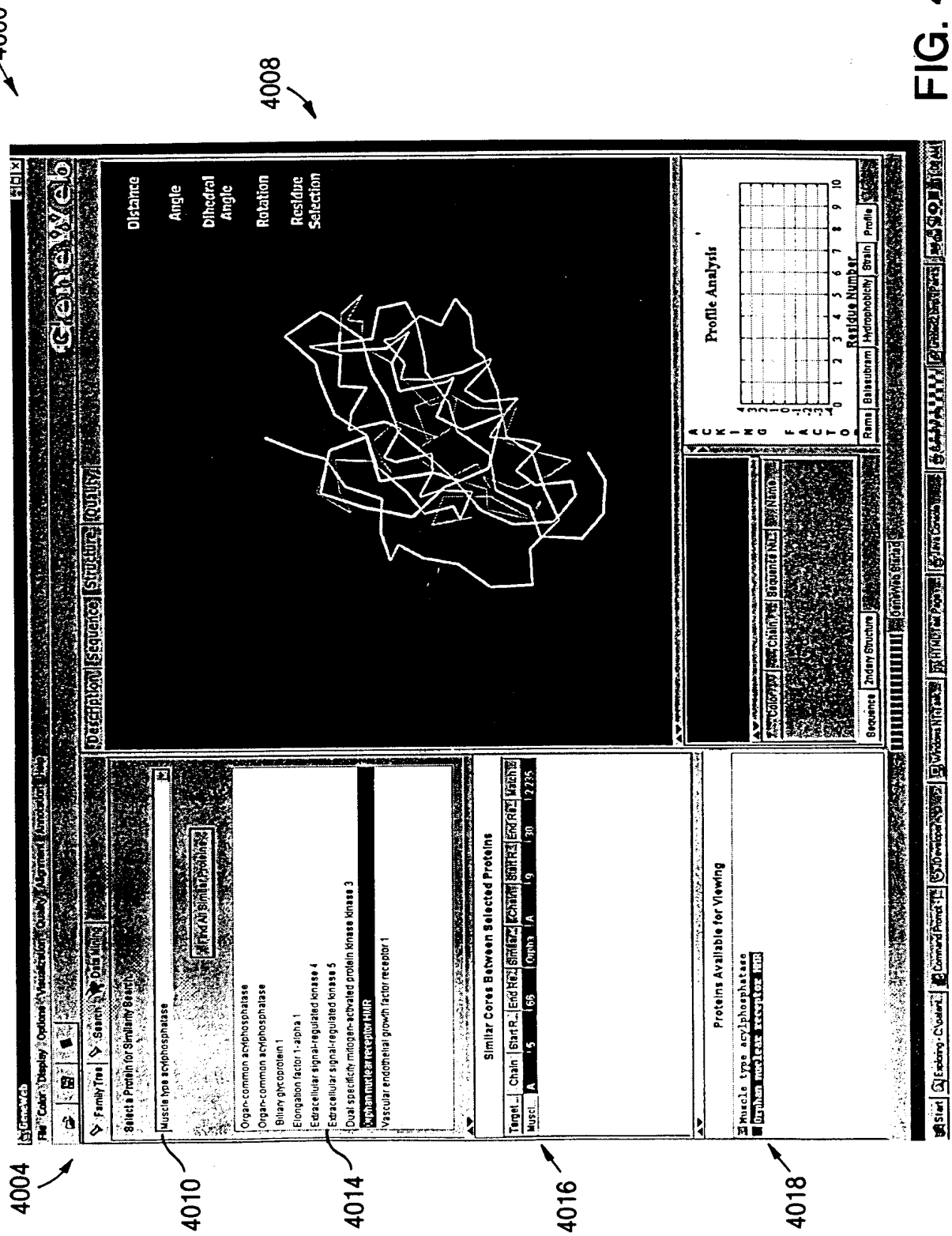

4000
4008
4004
4010
4014
4016
4018
GeneWeb
Distance
Angle
Dihedral Angle
Rotation
Residue Selection
Profile Analysis
Residue Number
Similar Cores Between Selected Proteins
Proteins Available for Viewing
Muscle type acylphosphatase
Human nuclear receptor HNR
Organ-common acylphosphatase
Organ-common acylphosphatase
Biliary glycoprotein 1
Elongation factor 1-alpha 1
Extracellular signal-regulated kinase 4
Extracellular signal-regulated kinase 5
Dual specificity mitogen-activated protein kinase kinase 3
Human nuclear receptor HNR
Vascular endothelial growth factor receptor 1

FIG. 40

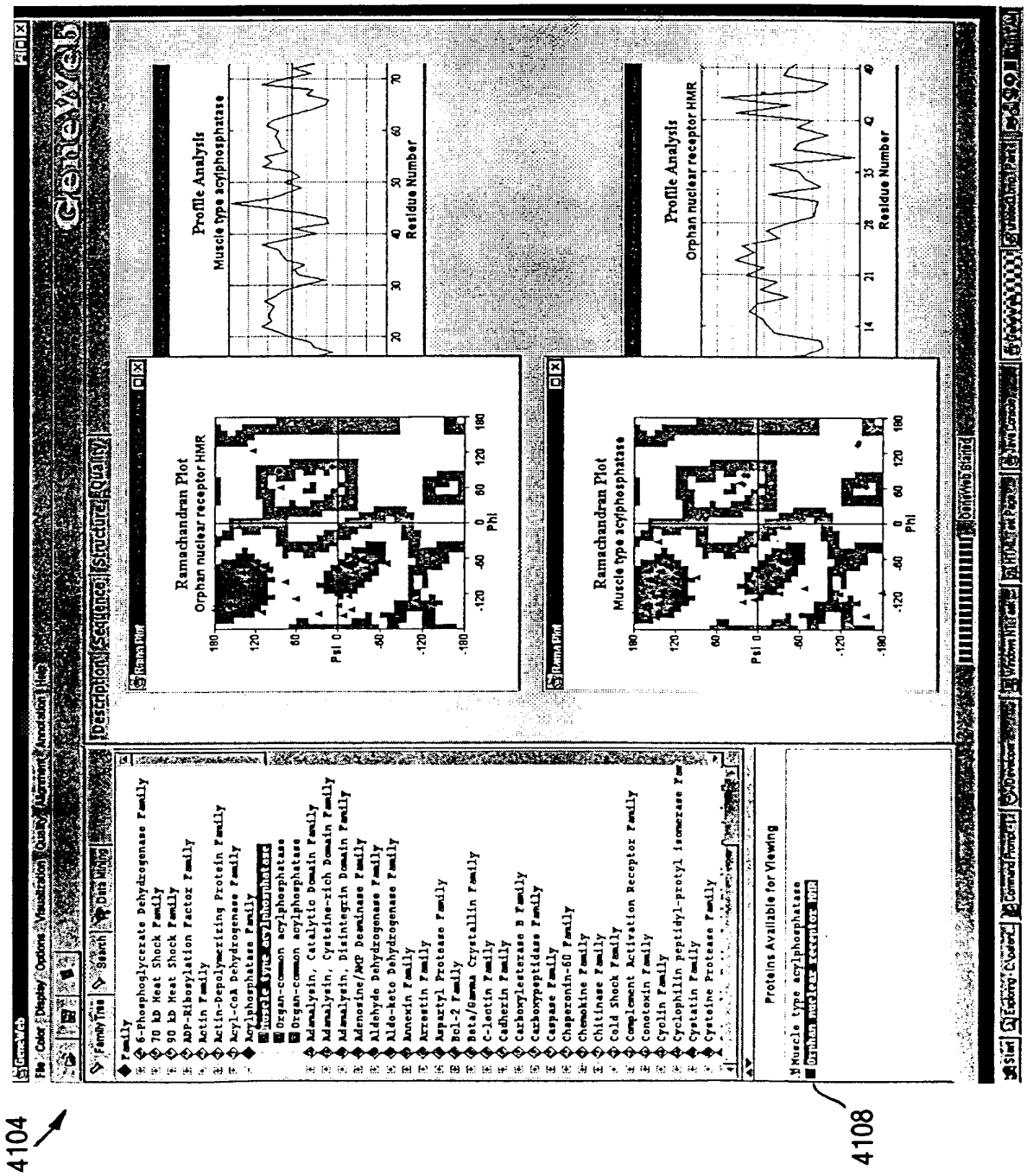

4100
4108
4104
4108
GeneWeb
Description
Sequence
Structure
Quality
Ramachandran Plot
Orphan nuclear receptor HMR
Profile Analysis
Muscle type acylphosphatase
Ramachandran Plot
Muscle type acylphosphatase
Profile Analysis
Orphan nuclear receptor HMR
Residue Number
Phi
Psi
Proteins Available for Viewing
Muscle type acylphosphatase
Orphan nuclear receptor HMR

FIG. 41

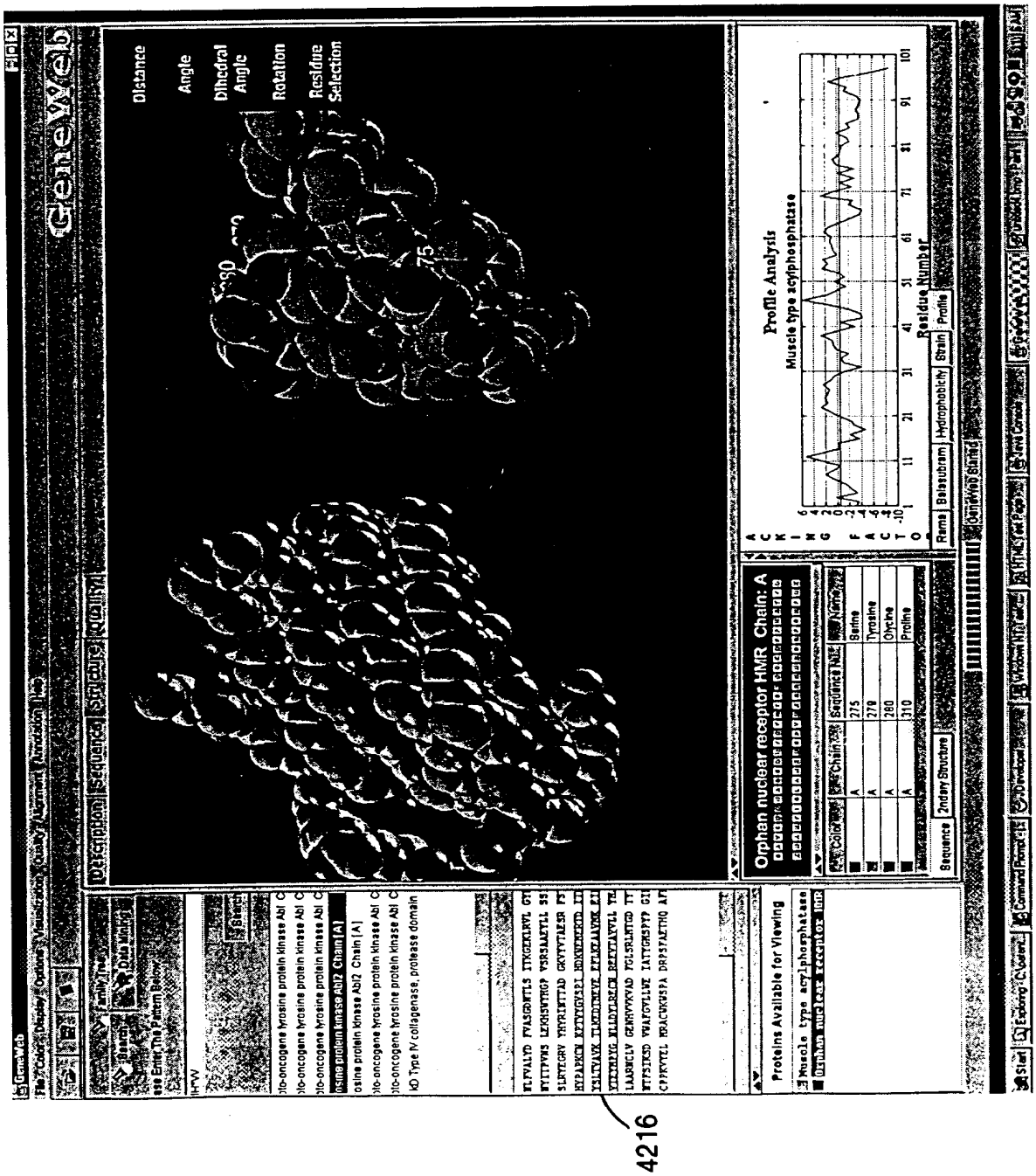

4200
4208
4216
GeneWeb
Distance
Angle
Dihedral Angle
Rotation
Residue Selection
Orphan nuclear receptor HNR Chain: A
Proteins Available for Viewing
Muscle type acylphosphatase
Profile Analysis
Muscle type acylphosphatase
Residue Number
Sequence
2ndary Structure
Rama
Ramachandran
Hydrophobicity
Strain
Profile

FIG. 42

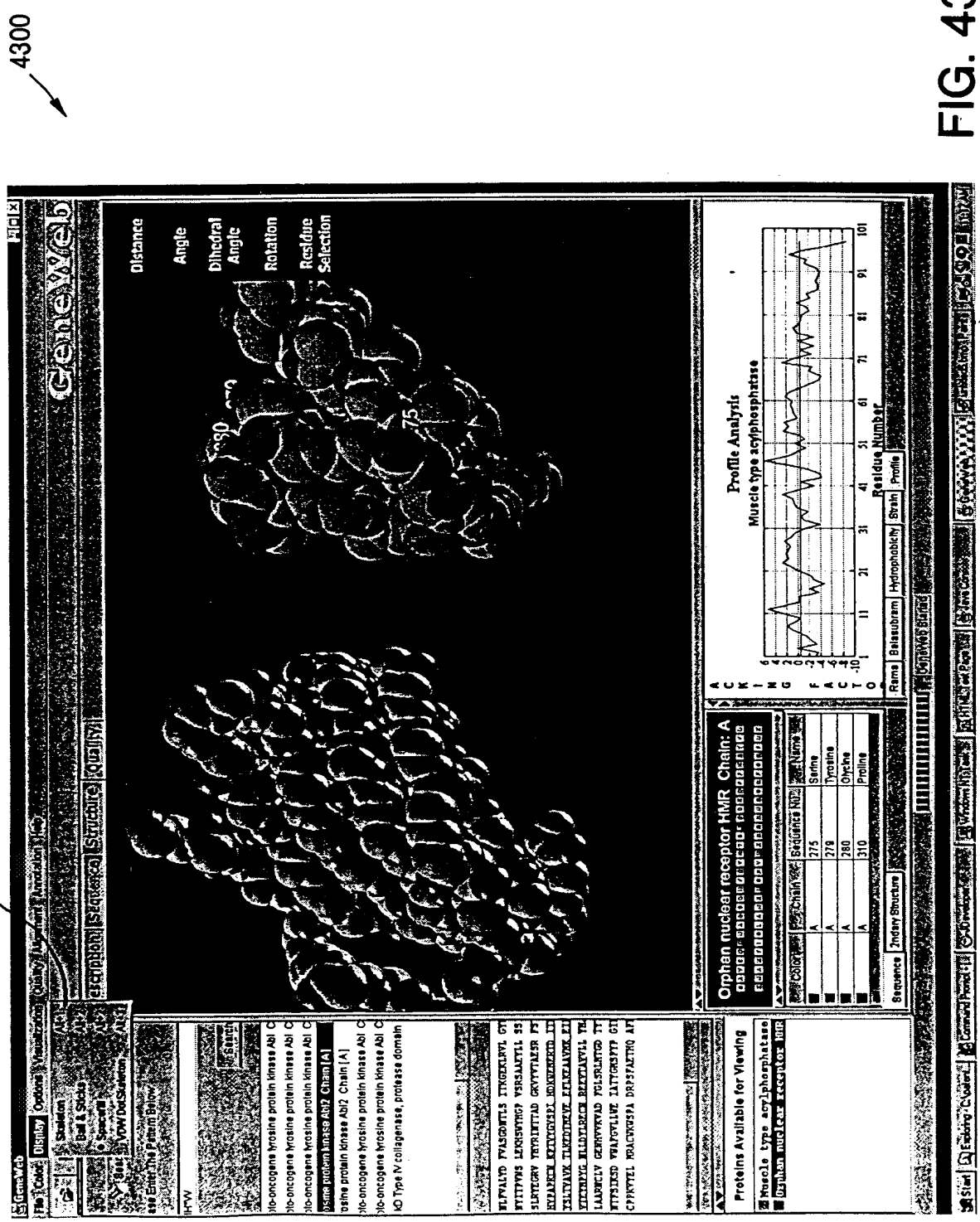

4300
4320
GeneWeb
Distance
Angle
Dihedral Angle
Rotation
Residue Selection
Profile Analysis
Muscle type acylphosphatase
Residue Number
Orphan nuclear receptor HMR Chain: A
Muscle type acylphosphatase
Orphan nuclear receptor HMR
Serine
Tyrosine
Glycine
Proline

FIG. 43

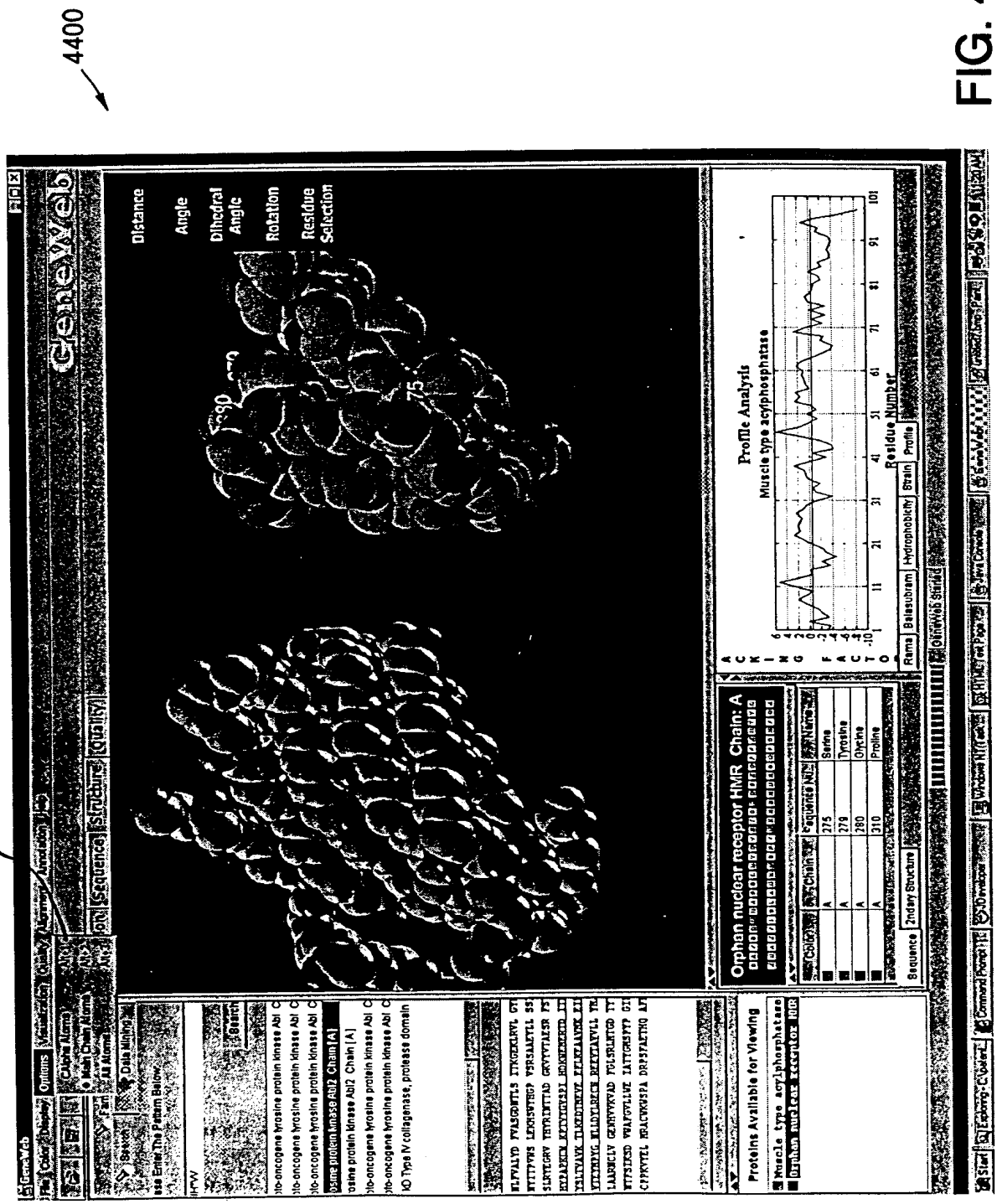

4400
4420
GeneWeb
Distance
Angle
Dihedral Angle
Rotation
Residue Selection
Profile Analysis
Muscle type acylphosphatase
Residue Number
Orphan nuclear receptor HNR Chain: A
Proteins Available for Viewing
Muscle type acylphosphatase
Orphan nuclear receptor HNR
Sequence
2ndary Structure
Rama
Ramachandran
Hydrophobicity
Strain
Profile

FIG. 44

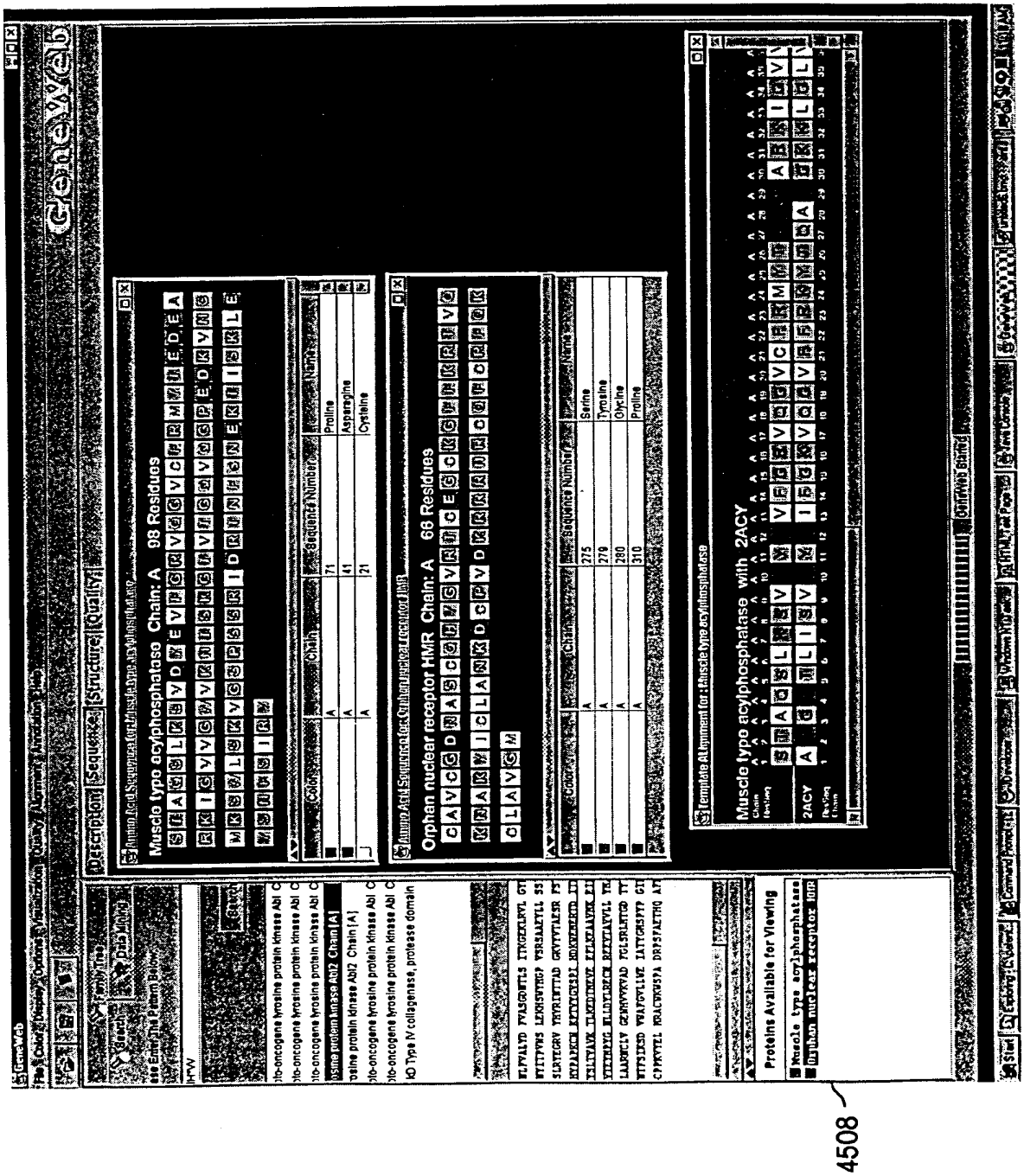
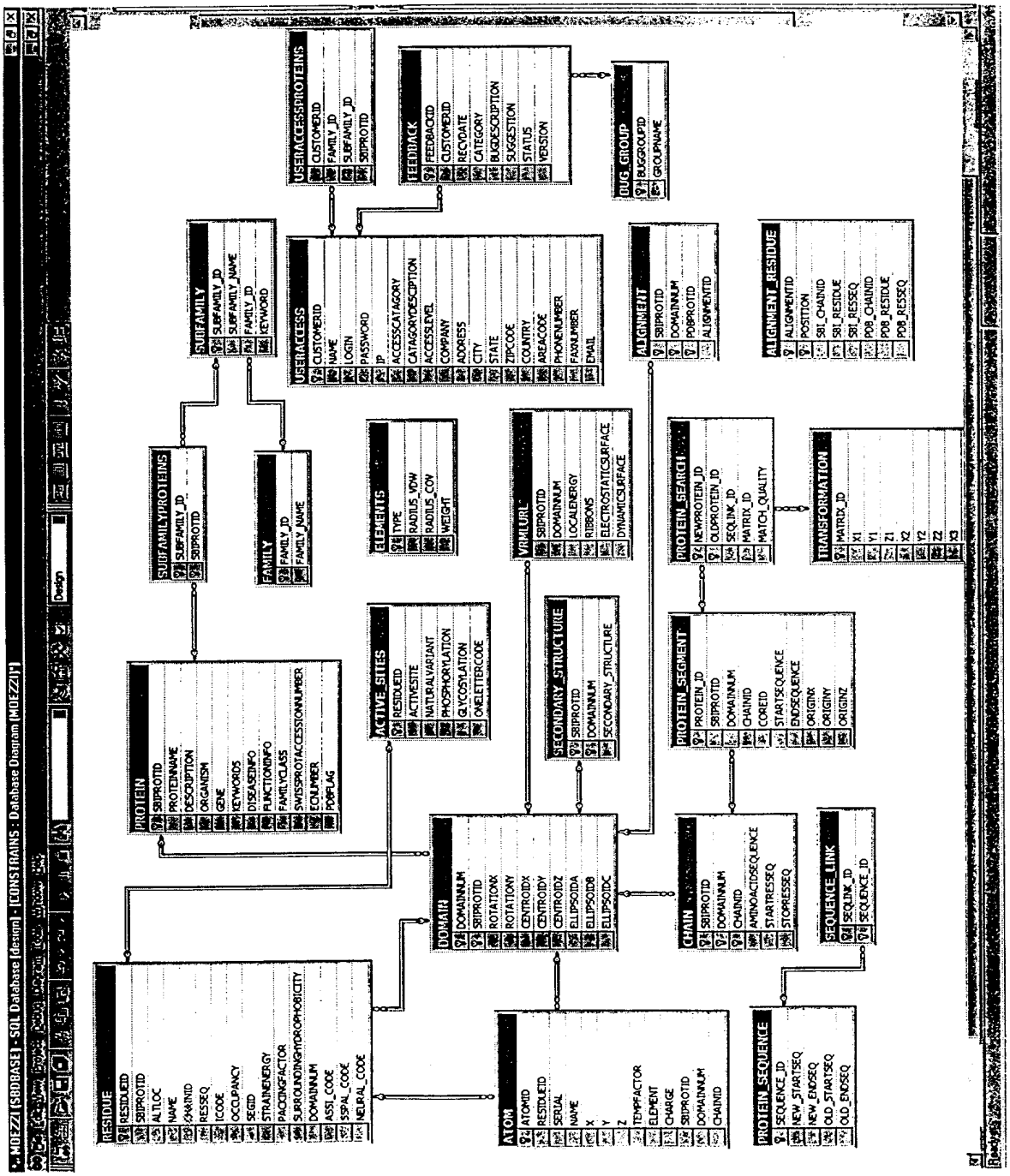

4500
4508
4508
GeneWeb
Description
Sequence
Structure
Quality
Muscle type acylphosphatase Chain: A 98 Residues
Orphan nuclear receptor HMR Chain: A 68 Residues
Muscle type acylphosphatase with 2ACY
Proteins Available for Viewing
Muscle type acylphosphatase
Orphan nuclear receptor HMR

FIG. 45

FIG. 46


: MOEZZI (SQLDBASE) - SQL Database Design - [ICONSTRAINS : Database Diagram (MOEZZI)]
Design
USERACCESSPROTEINS
CUSTOMERID
FAMILY_ID
SUBFAMILY_ID
SBJPROTID
FEEDBACK
FEEDBACKID
CUSTOMERID
RECVDATE
CATEGORY
BUGDESCRIPTION
SUGGESTION
STATUS
VERSION
BUG_GROUP
BUGGROUPID
GROUPNAME
SUBFAMILY
SUBFAMILY_ID
SUBFAMILY_NAME
FAMILY_ID
KEYWORD
USERACCESS
CUSTOMERID
NAME
LOGIN
PASSWORD
IP
ACCESSCATAGORY
CATAGORYDESCRIPTION
ACCESSLEVEL
COMPANY
ADDRESS
CITY
STATE
ZIPCODE
COUNTRY
AREACODE
PHONENUMBER
FAXNUMBER
EMAIL
ALIGNMENT
SBJPROTID
DOMAINNUM
PDBPROTID
ALIGNMENTID
ALIGNMENT_RESIDUE
ALIGNMENTID
POSITION
SBJ_CHAINID
SBJ_RESIDUE
SBJ_RESSEQ
PDB_CHAINID
PDB_RESIDUE
PDB_RESSEQ
SUBFAMILYPROTEINS
SUBFAMILY_ID
SBJPROTID
FAMILY
FAMILY_ID
FAMILY_NAME
ELEMENTS
TYPE
RADIUS_VDW
RADIUS_COV
WEIGHT
VRMLURL
SBJPROTID
DOMAINNUM
LOCALENERGY
RIBBONS
ELECTROSTATICSURFACE
DYNAMICSURFACE
PROTEIN_SEARCH
NEWPROTEIN_ID
OLDPROTEIN_ID
SEQLINK_ID
MATRIX_ID
MATCH_QUALITY
TRANSFORMATION
MATRIX_ID
X1
Y1
Z1
X2
Y2
Z2
X3
PROTEIN
SBJPROTID
PROTEINNAME
DESCRIPTION
ORGANISM
GENE
KEYWORDS
DISEASEINFO
FUNCTIONINFO
FAMILYCLASS
SWISSPROTACCESSIONNUMBER
EC_NUMBER
PDBFLAG
ACTIVE_SITES
RESIDUEID
ACTIVESITE
NATURALVARIANT
PHOSPHORYLATION
GLYCOSYLATION
ONELETTERCODE
SECONDARY_STRUCTURE
SBJPROTID
DOMAINNUM
SECONDARY_STRUCTURE
PROTEIN_SEGMENT
PROTEIN_ID
SBJPROTID
DOMAINNUM
CHAINID
COREID
STARTSEQUENCE
ENDSEQUENCE
ORIGINX
ORIGINY
ORIGINZ
RESIDUE
RESIDUEID
SBJPROTID
ALTLOC
NAME
CHAINID
RESSEQ
ICODE
OCCUPANCY
SEGID
STRAINENERGY
PACKINGFACTOR
SURROUNDINGHYDROPHOBICITY
DOMAINNUM
ASSI_CODE
SSPAL_CODE
NEURAL_CODE
DOMAIN
DOMAINNUM
SBJPROTID
ROTATIONX
ROTATIONY
CENTROIDX
CENTROIDY
CENTROIDZ
ELLIPSOIDA
ELLIPSOIDB
ELLIPSOIDC
CHAIN
SBJPROTID
DOMAINNUM
CHAINID
AMINOACIDSEQUENCE
STARTRESSEQ
STOPRESSEQ
SEQUENCE_LINK
SEQLINK_ID
SEQUENCE_ID
ATOM
ATOMID
RESIDUEID
SERIAL
NAME
X
Y
Z
TEMPFACTOR
ELEMENT
CHARGE
SBJPROTID
DOMAINNUM
CHAINID
PROTEIN_SEQUENCE
SEQUENCE_ID
NEW_STARTSEQ
NEW_ENDSEQ
OLD_STARTSEQ
OLD_ENDSEQ